



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: <b>XXXX XXX/XXXXX</b>	<b>A2</b>	(11) International Publication Number: <b>WO 00/04196</b>
		(43) International Publication Date: 27 January 2000 (27.01.2000)
(21) International Application Number: <b>PCT/US99/15860</b>	<b>Published</b>	
(22) International Filing Date: 13 July 1999 (13.07.1999)		
(30) Priority Data: 09/165,794 02 October 1998 (02.10.1998) US 60/092,697 14 July 1998 (14.07.1998) US		
(60) Parent Application or Grant CISTEM MOLECULAR CORPORATION [/]; ( ). KAUFFMAN, Stuart, A. [/]; ( ). BALLIVET, Marc [/]; ( ). KAUFFMAN, Stuart, A. [/]; ( ). BALLIVET, Marc [/]; ( ). GAY, David, A. ; ( ).		
(54) Title: CIS ACTING NUCLEIC ACID ELEMENTS AND METHODS OF USE (54) Titre: ELEMENTS D'ACIDES NUCLEIQUES A ACTIVITE CIS ET METHODES D'UTILISATION		
(57) Abstract <p>The invention provides a method of identifying nucleic acid molecules that contain cis acting nucleic acid elements. Also provided is a method of isolating nucleic acid binding factors. The invention also provides methods of identifying compounds that are cis acting nucleic acid element analogs, compounds that are nucleic acid binding factor analogs, compounds that selectively bind cis acting nucleic acid elements and compounds that selectively displace binding between a nucleic acid binding factor and a cis acting nucleic acid element or between nucleic acid binding factors. Also provided is a method of determining a binding state of a nucleic acid. Pluralities of isolated nucleic acid molecules containing cis acting nucleic acid elements, of isolated cis acting nucleic acid elements and of isolated nucleic acid binding factors are also provided. The invention further provides methods of treating pathological conditions using molecules of the invention to alter genetic activities of nucleic acids involved in pathological conditions.</p>		
(57) Abrégé <p>L'invention concerne une technique d'identification de molécules d'acides nucléiques contenant des éléments d'acides nucléiques à activité cis. L'invention concerne également une méthode permettant d'isoler des facteurs de liaison d'acides nucléiques, ainsi que des méthodes d'identification de composés constituant des analogues des éléments d'acides nucléiques à activité cis, de composés constituant des analogues des facteurs de liaison des acides nucléiques, de composés liant sélectivement les éléments d'acides nucléiques à activité cis et de composés déplaçant sélectivement la liaison entre un facteur de liaison d'acides nucléiques et un élément d'acide nucléique à activité cis ou entre plusieurs facteurs de liaison d'acides nucléiques. De surcroît, l'invention concerne d'une part, une méthode permettant de déterminer l'état de liaison d'un acide nucléique, et d'autre part plusieurs molécules d'acides nucléiques contenant des éléments d'acides nucléiques à activité cis, plusieurs éléments d'acides nucléiques à activité cis isolés, et plusieurs facteurs de liaison d'acides nucléiques isolés. L'invention concerne enfin des méthodes de traitement d'états pathologiques utilisant les molécules de l'invention pour modifier l'activité génétique des acides nucléiques impliqués dans ces états pathologiques.</p>		

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup>:</b> <b>C12Q 1 /68</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/04196</b> <b>(43) International Publication Date:</b> 27 January 2000 (27.01.00)
<b>(21) International Application Number:</b> PCT/US99/15860 <b>(22) International Filing Date:</b> 13 July 1999 (13.07.99) <b>(30) Priority Data:</b> 60/092.697 14 July 1998 (14.07.98) US 09/165.794 2 October 1998 (02.10.98) US <b>(71) Applicant (for all designated States except US):</b> CISTEM MOLECULAR CORPORATION [US/US]; 5310 Eastgate Mail, San Diego, CA 92064 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> KAUFFMAN, Stuart. A. [-/US]; 15 Montecito, Santa Fe, NM 87501 (US). BALLIVET, Marc [-/CH]; 15, rue Muzy, CH-1207 Geneva (CH). <b>(74) Agents:</b> GAY, David, A. et al.; Campbell & Flores LLP, Suite 700, 4370 La Jolla Village Drive, San Diego, CA 92122 (US).		<b>(81) Designated States:</b> AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> CIS ACTING NUCLEIC ACID ELEMENTS AND METHODS OF USE		
<b>(57) Abstract</b>  The invention provides a method of identifying nucleic acid molecules that contain cis acting nucleic acid elements. Also provided is a method of isolating nucleic acid binding factors. The invention also provides methods of identifying compounds that are cis acting nucleic acid element analogs, compounds that are nucleic acid binding factor analogs, compounds that selectively bind cis acting nucleic acid elements and compounds that selectively displace binding between a nucleic acid binding factor and a cis acting nucleic acid element or between nucleic acid binding factors. Also provided is a method of determining a binding state of a nucleic acid. Pluralities of isolated nucleic acid molecules containing cis acting nucleic acid elements, of isolated cis acting nucleic acid elements and of isolated nucleic acid binding factors are also provided. The invention further provides methods of treating pathological conditions using molecules of the invention to alter genetic activities of nucleic acids involved in pathological conditions.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GR	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**Descripti n**

**5**

**10**

**15**

**20**

**25**

**30**

**35**

**40**

**45**

**50**

**55**

CIS ACTING NUCLEIC ACID ELEMENTS AND METHODS OF USEBACKGROUND OF THE INVENTION

This invention relates to the identification and use of cis acting nucleic acid elements that bind to nucleic acid binding factors to regulate genetic activities of nucleic acids.

All living creatures store information in nucleic acid molecules called DNA or RNA that encode structural and regulatory proteins. The collective behavior of nucleic acids and proteins constitutes and controls normal cell and organismal life cycles. Nucleic acids and proteins also act as causative agents in, or response factors to, pathological conditions.

Transcription of DNA into RNA, translation of RNA into proteins and other genetic events such as nucleic acid synthesis, sorting, processing, repair and degradation, are regulated by a variety of specialized nucleic acid binding factors. Nucleic acid binding factors bind to specific sequences present on the nucleic acid molecules they regulate, called cis acting nucleic acid elements. These nucleic acid binding factors, bound to their specific cis acting nucleic acid elements, are able to interact with other cellular factors to modulate specific genetic events. The binding of a nucleic acid binding factor to a cis acting nucleic acid element, or its ability to interact with other factors that mediate genetic events, or both, can be regulated in response to signals transmitted into the cell from the cell exterior.

As an example, regulatory proteins called "transcription factors" bind to cis acting nucleic acid

elements on genomic DNA at sites known as "promoters" and "enhancers" present at variable distances from the site of initiation of transcription of the genes they regulate. The enhancer sequences and adjacent nucleic acid sequences, together with their bound transcription factors, are able to bend to contact the transcriptional complex bound to the promoter. Such contact can either enhance or reduce expression of the regulated gene.

The human genome, which stores the genetic information of a human cell as DNA, is estimated to contain about 100,000 genes. Each of these genes and the RNAs they encode is likely to have multiple cis acting nucleic acid elements that bind to corresponding nucleic acid binding factors to regulate gene expression. These cis acting nucleic acid elements, and the factors that bind them, are potential targets for therapeutic drugs that could be used to modulate gene expression. Determining which cis acting nucleic acid elements are bound under different conditions can also be used to characterize and monitor the genetic responses of a cell under normal, pathological or experimental conditions.

Current methods of identifying cis acting nucleic acid elements have several disadvantages. Most of these methods require prior identification of either the nucleic acid that is regulated, or the corresponding regulatory nucleic acid binding factor, or both. For example, once a nucleic acid has been identified, adjacent sequences, which are predicted to contain cis acting nucleic acid elements, can be isolated and subsequences therefrom are tested for cis activities. Alternatively, once a nucleic acid binding factor has been isolated, the sequences to which it binds can be identified. Other methods, which are limited to

identifying transcriptional enhancer elements, involve cloning random nucleic acid sequences upstream of a reporter gene and observing expression of the reporter gene product.

At present, however, there is no broadly applicable method to identify cis acting nucleic acid elements without prior identification of the regulated nucleic acid or of the regulatory nucleic acid binding factor. There is also no rapid and efficient method to simultaneously identify a plurality of cis acting nucleic acid elements.

Thus, there exists a need for a method of rapidly and efficiently identifying cis acting nucleic acid elements. The present invention satisfies this need and provides related advantages as well.

#### SUMMARY OF THE INVENTION

The invention provides a method of identifying nucleic acids containing cis acting nucleic acid elements. The method consists of contacting a diverse population of nucleic acid binding factors with a diverse population of isolated nucleic acid molecules under conditions that allow the nucleic acid binding factors to selectively bind the nucleic acids. The nucleic acids that bind the nucleic acid binding factors are identified and are characterized as nucleic acids containing cis acting nucleic acid elements. The method simultaneously provides for the isolation of nucleic acid binding factors that selectively bind the isolated nucleic acid molecules.

The invention also provides methods of identifying compounds that are cis acting nucleic acid element

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

analog, compounds that are nucleic acid binding factor analogs, and compounds that selectively bind cis acting nucleic acid elements. The invention further provides methods to identify compounds that selectively displace binding between a nucleic acid binding factor and a cis acting nucleic acid element or between nucleic acid binding factors.

The invention further provides a plurality of isolated nucleic acid molecules that each contain one or more cis acting nucleic acid elements. Also provided is a plurality of isolated cis acting nucleic acid element analogs. The isolated nucleic acid molecules containing cis acting nucleic acid elements and the isolated cis acting nucleic acid element analogs in the pluralities can be bound to nucleic acid binding factors. A plurality of isolated nucleic acid binding factors is also provided.

The invention also provides a method of determining a binding state of a nucleic acid. The method consists of contacting a nucleic acid with a plurality of isolated cis acting nucleic acid elements under conditions that allow nucleic acid binding factors bound to the nucleic acid to bind to the isolated cis acting nucleic acid elements. The isolated cis acting nucleic acid elements that bind the nucleic acid binding factors are identified and characterize the binding state of the nucleic acid.

The invention further provides a method of treating a pathological condition in an individual. The method consists of administering to the individual an effective amount of a therapeutic agent that selectively alters the ability of a cis acting nucleic acid element



5 to regulate a genetic activity of a nucleic acid involved  
in the pathological condition. Also provided is a method  
10 of treating a pathological condition in an individual by  
contacting a cell of the individual with an effective  
5 amount of a targeting construct that includes a cis  
acting nucleic acid element and targeting sequences. The  
targeting construct is taken up by the cell and inserted  
15 by homologous recombination into a nucleic acid involved  
in the pathological condition so as to alter a genetic  
10 activity of the nucleic acid.

20 DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the  
25 identification and use of cis acting nucleic acid  
elements.

15 Cis acting nucleic acid elements and the binding  
factors that selectively bind such elements regulate the  
30 genetic circuitry that controls all aspects of cell and  
organismal growth and development. Cis acting nucleic  
acid elements regulate genetic activities that underlie  
35 growth and development, including, for example,  
20 replication of nucleic acids and expression of both  
nucleic acids and proteins. Therefore, cis acting  
nucleic acid elements and their corresponding nucleic  
40 acid binding factors are targets for therapeutic agents  
25 that modulate cell or tissue growth, development,  
pathogenesis, regeneration or repair by altering,  
enhancing or reducing the genetic activity of the nucleic  
45 acids they regulate.

Compounds that selectively bind cis acting nucleic  
30 acid elements, that selectively bind nucleic acid binding  
factors, or that selectively displace binding of a cis  
50

5 acting nucleic acid element to its binding factor, are  
all potential therapeutic agents that can modulate a  
10 genetic activity of a nucleic acid regulated by the cis  
acting nucleic acid element. Furthermore, isolated cis  
5 acting nucleic acid elements and the corresponding  
nucleic acid binding factors can themselves be used as  
therapeutic agents to selectively modulate a genetic  
15 activity. Cis acting nucleic acid elements can also be  
used to identify and isolate a nucleic acid or group of  
10 nucleic acids that are modulated by the cis acting  
nucleic acid elements, such as a gene or a family of  
20 genes involved in a particular disease or that regulate a  
particular stage of development.

25 In one embodiment, the invention provides methods  
15 of identifying cis acting nucleic acid elements. The  
methods are advantageous in allowing rapid and efficient  
identification of cis acting nucleic acid elements  
30 without prior knowledge of the nucleic acid sequences  
they regulate or of the corresponding nucleic acid  
20 binding factors that bind the cis acting elements. The  
methods provide a means of simultaneously identifying cis  
35 acting nucleic acid elements that modulate a genetic  
activity of a plurality of nucleic acids. Cis acting  
nucleic acid elements can be used as therapeutic agents  
25 or to screen for therapeutic agents, as well as to  
40 diagnose disease.

45 In another embodiment, the invention provides  
methods for identifying nucleic acid binding factors that  
bind to cis acting nucleic acid elements without prior  
30 knowledge of either the cis acting nucleic acid elements  
they bind or the nucleic acid sequences they regulate.  
The methods are advantageous in providing a means of  
50 simultaneously identifying nucleic acid binding factors

5 that modulate a genetic activity of a plurality of  
nucleic acids. Nucleic acid binding factors can be used  
10 as therapeutic agents or to screen for therapeutic agents  
that selectively target a nucleic acid or group of  
5 nucleic acids.

15 In yet another embodiment, the invention provides  
methods of identifying compounds that are analogs of cis  
acting nucleic acid elements or of nucleic acid binding  
factors, or that displace binding of cis acting nucleic  
20 acid elements to nucleic acid binding factors. The  
methods are advantageous in that they provide a rapid and  
efficient means of screening for compounds that can be  
used as therapeutic agents to modulate a genetic activity  
25 of a nucleic acid or group of nucleic acids involved in  
15 disease.

30 In another embodiment, the invention is directed  
to a method of determining the binding state of one or a  
plurality of nucleic acids. The binding of a nucleic  
acid binding factor to a cis acting nucleic acid element  
20 is generally required for its regulatory activity.  
Therefore, the binding state of a nucleic acid or a  
35 plurality of nucleic acids is a means of characterizing  
the activation state of the nucleic acid or plurality of  
nucleic acids. Such a characterization can be used for a  
40 25 variety of purposes such as, for example, diagnosing  
pathological conditions or monitoring the efficacy of  
therapeutic procedures.

45 As used herein, the term "cis acting nucleic acid  
element" refers to a single-stranded or double-stranded  
30 RNA or DNA sequence that can be selectively bound by  
nucleic acid binding factors to regulate one or more  
50 genetic activities of a nucleic acid sequence present on

5 the same molecule. Cis acting nucleic acid elements are  
present in all organisms, including prokaryotes,  
10 eukaryotes and viruses. For example, cis acting nucleic  
acid elements are present in yeast, animals, plants,  
5 bacteria and viruses.

15 Cis acting DNA elements are found in a variety of  
different types of DNA including, for example, genomic,  
mitochondrial and chloroplast DNA. Cis acting DNA  
elements are also located at a variety of locations on  
20 chromosomes. For example, cis acting DNA elements are  
located at diverse locations within chromosomes, such as  
within transcription units or at the domain boundaries of  
transcriptional units, as well as at the centromeres,  
25 kinetochores and telomeres of chromosomes. Cis acting  
DNA elements can regulate a variety of genetic activities  
including, for example, enhancing, attenuating or  
repressing transcription of a structural or regulatory  
30 gene or operon. A cis acting DNA element can also  
regulate, for example, replication, repair, packaging,  
20 modification, restriction or degradation of a DNA  
sequence.

35 Cis acting DNA elements also include nucleic acid  
elements that modulate the assembly or structural  
integrity of DNA. A specific example of a cis acting DNA  
40 25 element that modulates the assembly or structural  
integrity of DNA is a boundary element that selectively  
binds to scaffold proteins and serves to define  
transcriptional domains of chromatin. Additionally, cis  
45 acting DNA elements are present at kinetochores,  
30 centromeres or telomeres of chromosomes and modulate the  
assembly and structural integrity of DNA.

50

55

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

Cis acting RNA elements are also found in a variety of different types of RNAs including, for example, messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), heterogeneous nuclear RNA (hnRNA), small nuclear or small cytoplasmic RNA (snRNA or scRNA) and viral RNA. Cis acting RNA elements can regulate a variety of genetic activities including, for example, RNA translation, replication, splicing, editing, intracellular transport, localization, degradation and reverse transcription.

20  
25  
30  
35  
40  
45  
50  
55

The types of cis acting nucleic acid elements present in nucleic acids vary depending on the cell and nucleic acid type. For example, transcription of eukaryotic DNA involves a variety of cis acting nucleic acid elements such as promoter elements, enhancer elements and response elements. Certain of these cis acting nucleic acid elements, for example, TATA boxes, are found in a majority of genes. Other cis acting nucleic acid elements, for example, hormone response elements, are characteristic of genes that are coordinately regulated. Some cis acting nucleic acid elements bind to nucleic acid binding factors in a tissue-specific or temporal manner, whereas others are constitutively bound by nucleic acid binding factors. Although individual cis acting nucleic acid elements can be involved in the regulation of many different nucleic acids, a particular combination of cis acting nucleic acid elements can be specific for one or only a limited number nucleic acids.

30  
35  
40  
45  
50  
55

A cis acting nucleic acid element can be localized within the nucleic acid sequence it regulates, or upstream or downstream thereof. A cis acting nucleic acid element can be a contiguous nucleic acid sequence,

5 or a multi-partite sequence. For example, a nucleic acid  
binding factor or complex of factors can bind to a  
continuous cis acting nucleic acid element or to two or  
10 more discontinuous nucleic acid sequences that are in  
5 close proximity due to folding or looping of the  
polynucleotide, that together form a nucleic acid  
element. A cis acting nucleic acid element is generally  
15 from about 4 to about 100 nucleotides in length, and is  
more typically from about 6 to about 25 nucleotides in  
10 length.

20 The methods of the invention are applicable to the  
identification and use of cis acting nucleic acid  
elements of a wide variety of nucleic acid types and  
sizes, and from any organism. The methods of the  
25 invention also allow the identification and use of cis  
15 acting nucleic acid elements or combinations of cis  
acting nucleic acid elements that modulate any regulatory  
or structural genetic activity, and that modulate any  
30 subset of nucleic acids that is of interest.

20 As used herein, the term "selective binding" or  
35 "selectively binds," when used in connection with binding  
between a cis acting nucleic acid element and either a  
nucleic acid binding factor or a compound, refers to  
binding with substantially higher affinity to a nucleic  
40 25 acid having a sequence that is substantially similar to  
the sequence of a particular cis acting nucleic acid  
element than to a nucleic acid that lacks substantial  
similarity to the sequence of a particular cis acting  
45 nucleic acid element. The degree or extent of nucleic  
30 acid sequence similarity required for selective binding  
of a nucleic acid binding factor or compound to a  
particular cis acting nucleic acid element depends on,  
50 for example, the length and sequence composition of the

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

cis acting nucleic acid element and the nature of the binding interaction. Such selective binding can be determined either qualitatively or quantitatively by known methods, such as by competition with nucleic acids of similar or different sequences to the cis acting nucleic acid element.

15  
20  
25  
30  
35  
40  
45  
50  
55

Selective binding between a nucleic acid binding factor and a compound refers to binding with substantially higher affinity to a substantially similar binding factor or compound than to an unrelated binding factor or compound. Selective binding between a nucleic acid binding factor and a compound can similarly be determined by, for example, competition for, or displacement of, binding with substantially similar binding factors and compounds, as compared with binding factors and compounds that lack substantially similarity. Selective binding between a nucleic acid binding factor and a compound that is a cis acting nucleic acid element analog can further be determined by an ability of a nucleic acid containing a sequence that is substantially similar to a cis acting nucleic acid element to compete for binding with the analog compound for the binding factor, such that the analog compound is selectively displaced.

40  
45  
50  
55

25  
30

As used herein, the term "diverse population of isolated nucleic acid molecules" refers to a composition comprising a plurality of different isolated polynucleotide nucleic acid molecules that potentially contain cis acting nucleic acid elements. The diverse population of nucleic acids used in the methods of the invention can be of a variety of different types, structures and topology. The choice of nucleic acid type, structure and topology will depend on the need and

desired result. For example, the diverse populations of nucleic acids of the invention can include double-stranded or single-stranded DNA or RNA, as well as linear, circular or branched nucleic acid molecules.

The term "isolated," when used in reference to isolated nucleic acid molecules, is intended to mean that the nucleic acid molecules are present in a form or state different from how they are found in nature. Similarly, the term "isolated," when used in reference to isolated nucleic acid binding factors, is intended to mean that the nucleic acid binding factors are present in a form or state different from how they are found in nature. For example, the isolated molecules can be different than populations found in nature in that they are substantially purified and therefore are free of molecules other than nucleic acids or other than nucleic acid binding factors. Such molecules can also be different than molecules found in nature in that they are, for example, produced or expressed by recombinant means or synthesized by chemical means. Such recombinantly or chemically produced molecules therefore do not contain some or many of the normal cellular components as they are found in nature or as they are isolated from natural sources and can also differ in multiplicity or homogeneity from populations of molecules found in nature. Furthermore, such molecules can also be different than molecules found in nature in that they are bound or immobilized, with or without cellular constituents, on a filter or solid support. Isolated molecules can also be different from the state or form found in nature in that they are detectably labeled or contain non-native nucleic acid sequences.



5 A population of different isolated nucleic acid  
molecules can be prepared, or obtained, that is of any  
diversity that is appropriate for a particular  
10 application of a method of the invention. A population  
5 of nucleic acids of low diversity can contain, for  
example, 2, 3, 4, 5, 6, 7, 8, 9, between about 10 and 20,  
between about 21 and 80, or between about 81 and 200  
15 different nucleic acid molecules. For certain  
applications of the method, it may be preferable to begin  
10 with a population of nucleic acids of moderate diversity,  
containing, for example, between about 200 and  $10^3$ ,  
20 preferably greater than about  $10^4$ , more preferably greater  
than about  $10^5$  different nucleic acid molecules. If  
desired, using currently available methods, it is  
25 possible to synthesize a population of isolated nucleic  
acid molecules of high diversity, containing, for  
example, between about  $10^6$  and  $10^8$  different nucleic acid  
molecules, preferably between about  $10^8$  and  $10^{11}$  different  
30 nucleic acid molecules, most preferably about  $10^{13}$   
20 different nucleic acid molecules. As an example, a  
population that includes all possible molecules of  
between 5 and 20 nucleotides in length, including each of  
35 the four naturally occurring nucleotides at each  
position, would have approximately  $4^5 + 4^6 + 4^7 + \dots + 4^{20}$  or  
25 approximately  $10^{13}$  different nucleic acid molecules. Such  
a population of about  $10^{13}$  20 different nucleic acid  
40 molecules inherently includes all possible cis acting  
nucleic acid elements of up to about 20 nucleotides in  
length.

45 30 A diverse population of isolated nucleic acid  
molecules can be of completely random composition or of  
partially or completely known composition, so long as  
some nucleic acid sequences within the population are  
50 different. One skilled in the art would be able to

determine the extent of diversity and degree of randomness required for a particular application of the method.

A diverse population of isolated nucleic acid molecules includes nucleic acid molecules potentially containing cis acting nucleic acid elements. Depending on the application of the method, a diverse population of isolated nucleic acid molecules can include single-stranded or double-stranded RNA or DNA molecules, or any combination thereof. The isolated nucleic acid molecules in the diverse population can be from about 4 to about 1000 nucleotides in length and can include molecules of the same or of varying lengths. If desired, some or all of the isolated nucleic acid molecules can include, or be flanked at one or both ends by, known sequences, such as sequences homologous to oligonucleotide primers for the polymerase chain reaction (PCR), sequences containing restriction sites, or detectable sequences.

As used herein, the term "nucleic acid binding factor" is a factor that selectively binds a cis acting nucleic acid element to modulate a genetic activity of a nucleic acid or group of nucleic acids. Modulation can include, for example, enhancing, repressing or attenuating the regulation of a nucleic acid. Nucleic acid binding factors include, for example, transcription factors, replication factors, translation factors, restriction and modifying factors, structural and assembly factors, and other molecules involved in regulating one or more genetic activities of a nucleic acid sequence. Nucleic acid binding factors also include factors involved in the structural integrity of chromatin or chromosomes, such as, for example, scaffold proteins

5 and other factors that selectively bind to boundary  
elements, kinetochores, centromeres and telomeres.

10 A nucleic acid binding factor can interact  
covalently or non-covalently with other factors to form a  
5 complex that binds a cis acting nucleic acid element.  
The factors within such a binding complex are also  
15 included within the term "nucleic acid binding factor."  
Some nucleic acid binding factors within a complex of  
nucleic acid binding factors can contact a cis acting  
10 nucleic acid element directly. Other nucleic acid  
binding factors within a complex of nucleic acid binding  
20 factors do not contact a cis acting nucleic acid element  
directly, but can contact one or more other nucleic acid  
binding factors. Disrupting the interaction between two  
25 or more nucleic acid binding factors within a complex, or  
between nucleic acid binding factors and a cis acting  
nucleic acid element, will alter the ability of the cis  
30 acting nucleic acid element to modulate a genetic  
activity of the nucleic acid it regulates.

20 A nucleic acid binding factor can be a polypeptide  
or a polypeptide that is modified, for example, by  
35 phosphorylation or addition of one or more carbohydrates,  
nucleotides, nucleic acids, cofactors or lipids. A  
nucleic acid binding factor can also be a non-  
40 25 proteinaceous molecule, such as a lipid, carbohydrate or  
nucleic acid, or any combination thereof.

45 As used herein, the term "diverse population of  
nucleic acid binding factors" is intended to mean a  
composition containing a plurality of different nucleic  
30 acid binding factors. The greater the number of  
different factors within the population, the greater the  
diversity of the population. A population of nucleic  
50 acid binding factors can be of low diversity for certain

5 applications of the method. For example, a population of  
nucleic acid binding factors of low diversity can  
include, for example, 2, 3, 4, 5, 6, 7, 8, 9, between  
10 about 10 and 20, between about 21 and 50, or between  
5 about 51 and 100 different nucleic acid binding factors.  
A population of nucleic acid binding factors of higher  
diversity can include more than about 100, more than  
15 about  $10^3$ , or more than about  $10^4$  different nucleic acid  
binding factors. As with the diverse populations of  
10 isolated nucleic acid molecules, the members within a  
diverse population of nucleic acid binding factors can be  
20 known, unknown or partially known so long as some of the  
factors are different. One skilled in the art would be  
able to determine the size and extent of diversity in a  
15 population of nucleic acid binding factors required to  
25 practice a particular embodiment of the invention.

A diverse population of nucleic acid binding  
30 factors can be a population of nucleic acid binding  
factors that is bound to nucleic acids, or unbound. For  
20 example, a population of nucleic acid binding factors  
bound to nucleic acids can be a cellular nucleic acid  
35 preparation that contains nucleic acid binding factors.  
Such a preparation can be, for example, a chromatin  
preparation, a hnRNA preparation, an mRNA preparation, or  
25 other nucleic acid preparation that includes nucleic acid  
40 binding factors, depending on the type and function of  
cis acting nucleic acid elements and nucleic acid binding  
factors that are desired to be obtained. A population of  
unbound nucleic acid binding factors can be, for example,  
45 30 a population of nucleic acid binding factors eluted from  
a nucleic acid preparation, or a cellular extract or  
subset thereof.

5 As used herein, the term "diverse population of  
compounds" refers to a plurality of different molecules  
10 that potentially includes therapeutic compounds that can  
be used to selectively bind to cis acting nucleic acid  
5 elements, to nucleic acid binding factors, or to both.  
Therefore, a diverse population of compounds can include  
analogs of cis acting nucleic acid elements, analogs of  
15 nucleic acid binding factors, and molecules that  
selectively displace the binding between a cis acting  
10 nucleic acid element and its corresponding binding  
factor. Such compounds can be naturally occurring  
20 macromolecules, such as polypeptides, nucleic acids,  
carbohydrates or lipids. However, derivatives, analogs  
and mimetics of these macromolecules, as well as organic  
15 compounds, including polymers and small organic  
25 compounds, can also selectively bind a cis acting nucleic  
acid element or a nucleic acid binding factor.

30 The extent of diversity of a population of  
compounds required for a particular application of  
20 methods of the invention can be determined by those  
skilled in the art. Generally, the greater the  
35 diversity, the larger the likelihood of identifying a  
compound that binds a cis acting nucleic acid element or  
a nucleic acid binding factor, or that displaces binding  
25 between a cis acting nucleic acid element and a nucleic  
acid binding factor. A population of compounds of  
40 moderate diversity can readily be produced or obtained  
that contains greater than about  $10^5$  different compounds,  
more preferably greater than about  $10^7$  different  
45 30 compounds. A highly diverse population of compounds that  
contains greater than about  $10^9$ , preferably greater than  
about  $10^{11}$ , more preferably greater than about  $10^{13}$   
50 different compounds, can also be used in a method of the  
invention and can be readily produced or obtained. A

5 less diverse population of compounds can also be  
advantageous, for example, if the type of compounds that  
10 are likely to bind are known or can be predicted based  
on, for example, information about the sequence or  
5 structure of the cis acting nucleic acid element, the  
nucleic acid binding factor, or the binding interaction  
between them.

15 A diverse population of compounds can include, for  
example, naturally occurring nucleic acids and modified  
20 nucleic acids that contain non-naturally occurring  
nucleoside analogs or linkages. Such modifications can  
be advantageous, for example, for increasing resistance  
to chemical or enzymatic degradation. Various  
25 modifications that increase the stability of nucleic  
15 acids are known in the art and include, for example,  
phosphotioate linkages. Methods of producing diverse  
populations of natural and modified nucleic acids are  
30 known in the art.

A diverse population of compounds that potentially  
20 includes therapeutic agents that target cis acting  
35 nucleic acid elements or nucleic acid binding factors can  
also include libraries of peptides, carbohydrates or  
synthetic organic molecule. Peptide libraries can  
include, for example, diverse populations of chemically  
40 25 synthesized peptides and peptidomimetic molecules.  
Peptide libraries can also include populations of  
peptides generated by recombinant means, such as phage  
display or other recombinant methodologies by which a  
45 peptide is or can be associated with the nucleic acid  
30 which encodes it. Peptide and peptidomimetic libraries  
of high diversity can be obtained commercially or can be  
produced by methods known in the art. A diverse  
50 population of compounds that potentially includes

therapeutic agents that target cis acting nucleic acid elements or nucleic acid binding factors can be a carbohydrate-based combinatorial library, such as an oligosaccharide and glycoconjugate library. Diverse populations of small synthetic molecules, prepared by combinatorial chemistry methods, are also commercially available or can be produced by means known in the art. For example, a diverse population of organic molecules that share one or more common structural features but vary in reactive groups can be routinely produced. Any of these libraries of compounds, if desired, can be synthesized or immobilized onto a solid support or detectably tagged by methods known in the art to provide a means of detection.

As used herein, the term "binding state" refers to the condition or degree of binding of cis acting nucleic acids by nucleic acid binding factors. Modulation, including activation, repression and attenuation of the genetic properties of a nucleic acid by a cis acting nucleic acid element often requires binding of a nucleic acid binding factor to the cis acting nucleic acid element. Therefore, the binding state of a nucleic acid is a reflection or measurement of the type, degree, or extent of regulation of the nucleic acid.

Determination of a "binding state" can be either qualitative or quantitative. For certain applications, it may be sufficient to determine whether one or a plurality of nucleic acids is or is not bound by any nucleic acid binding factor or by a particular nucleic acid binding factor. For other applications, it may be desirable to determine to what degree or extent a nucleic acid is bound by a nucleic acid binding factor. For example, it may be desirable to determine the percentage

5 of nucleic acids that are bound by a nucleic acid binding  
factor, or to determine the affinity of a binding  
10 interaction. For certain determinations of the binding  
state, it may also be desirable to identify the nucleic  
5 acid binding factor that binds the nucleic acid.

15 Depending on the particular nucleic acids and  
isolated cis acting nucleic acid elements used in an  
application of the method, the term "binding state" can  
refer to, for example, the "transcriptional state," the  
20 "replication state," the "translational state" or other  
genetic properties of a nucleic acid. Furthermore, the  
term "binding state" can refer to a binding state of a  
single nucleic acid or group of nucleic acids. The term  
25 "binding state" can also refer to the binding state of a  
15 cell, group of cells, or tissue. For example, the term  
"binding state" can characterize the transcriptional  
activation state of a gene or a family of genes in a cell  
type of interest.  
30

The invention provides a method of identifying a  
20 nucleic acid containing a cis acting nucleic acid  
element. The method involves contacting a diverse  
35 population of nucleic acid binding factors with a diverse  
population of isolated nucleic acid molecules under  
conditions that allow nucleic acid binding factors to  
40 25 selectively bind the nucleic acids. The nucleic acids  
that selectively bind the nucleic acid binding factors  
are identified and are characterized as nucleic acids  
containing a cis acting nucleic acid element.  
45

As described previously, cis acting nucleic acid  
30 elements selectively bind nucleic acid binding factors  
and modulate one or more genetic activities of nearby  
50 nucleic acids. Any method of altering the interaction



5 between a cis acting nucleic acid element and a nucleic  
acid binding factor can be used to alter a genetic  
activity of the regulated nucleic acid. For example,  
10 selective binding between a cis acting nucleic acid  
5 element and a nucleic acid binding factor can be  
displaced by a molecule that selectively binds to either  
the cis acting nucleic acid element or the nucleic acid  
15 binding factor. Such a molecule can be, for example, a  
nucleic acid containing a cis acting nucleic acid  
20 element, a nucleic acid binding factor, or other  
compound. Similarly, selective binding between a  
compound and a nucleic acid binding factor can be  
selectively displaced by either a nucleic acid binding  
factor or a nucleic acid containing a cis acting nucleic  
25 acid element. Likewise, selective binding between a  
compound and a cis acting nucleic acid element can be  
selectively displaced by either a cis acting nucleic acid  
element or a nucleic acid binding factor. The molecules  
30 that are displaced and the molecules that effect the  
displacement, or any combination of these molecules, can  
be identified and isolated by a method of the invention.  
Therefore, by providing methods of distinguishing between  
35 nucleic acids that are bound by nucleic acid binding  
factors or other compounds, and nucleic acids that are  
unbound, the methods of the invention can be applied to  
25 the identification and isolation of cis acting nucleic  
acid elements, nucleic acid binding factors and compounds  
40 that bind either cis acting nucleic acid elements or  
nucleic acid binding factors.

45 30 The cis acting nucleic acid elements, nucleic acid  
binding factors and compounds identified by the methods  
of the invention can be used for therapeutic purposes to  
alter the activity of one or a plurality of nucleic acids  
50 involved, for example, in disease, development, tissue

5 repair or regeneration. The invention can be used with  
large, diverse populations of isolated nucleic acid  
10 molecules or nucleic acid binding factors, or smaller  
biased populations that contain, for example, nucleic  
5 acid sequences or nucleic acid binding factors that are  
known or predicted to be localized to a particular  
genomic region, or that are known or predicted to be  
15 indicative of a particular normal or pathological  
condition.

10  
A diverse population of isolated nucleic acid  
20 molecules can be produced or obtained by a variety of  
means known in the art. Both the diversity of the  
population and the type of nucleic acids will depend on  
15 the particular application of the method. Methods of  
25 producing a diverse population of isolated nucleic acid  
molecules are well known, and include, for example,  
biochemical and recombinant methods as well as by  
chemical synthesis. For example, a diverse population of  
30 isolated nucleic acid molecules can be obtained by  
20 cleaving an appropriate cellular or viral source of  
nucleic acids into smaller fragments by enzymatic,  
mechanical or chemical means. Fragments of approximately  
35 the desired size are isolated by fractionation methods  
25 known in the art, such as column chromatography or  
electrophoresis through a gel. As described previously,  
40 such fragments can be, for example, from about 4 to about  
1000 nucleotides in length.

45  
30 Subregions of the genome are particularly useful in  
applications where it is desirable to identify cis acting  
nucleic acid elements that regulate genes or gene  
families known or predicted to be involved in growth,  
development or pathogenesis. Therefore, a source of  
50 double-stranded DNA that can be fragmented to form a

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

diverse population of isolated nucleic acid molecules can be, for example, genomic DNA or a fragment therefrom, such as a chromosome or chromosomal arm, one or more DNA structural or transcriptional domains, or one or more genes. Methods of isolating such DNA preparations are known in the art. A source of single-stranded DNA can be, for example, any of the above double-stranded DNAs that either prior to or after fragmenting has been denatured by methods known in the art, including heating and alkali treatment. Similarly, sources of RNA, such as hnRNA, mRNA and viral RNA can be produced and fragmented or fractionated by means known in the art. If desired, known nucleic acid sequences can be attached to one or both ends of the isolated nucleic acid molecules.

15  
20  
25  
30  
35  
40  
45  
50  
55

A diverse population of isolated nucleic acid molecules of various lengths and sequence compositions can also be produced by synthetic means. For example, single-stranded DNA or RNA molecules can be synthesized using automatic nucleic acid synthesizers. Such molecules can include predetermined degenerate or random sequences at all or some positions. Methods of synthesis that result in random, degenerate or partially degenerate nucleic acid sequences are known in the art (see, for example, U.S. Patent No. 5,723,323, incorporated herein by reference). If desired, known nucleic acid sequences can be attached to one or both ends of the isolated nucleic acid molecules. Depending on the need, single-stranded nucleic acids can be rendered double-stranded and purified by means known in the art.

30  
35  
40  
45  
50  
55

The size of the diverse population of isolated nucleic acid molecules can vary depending on the need and desired efficiency for identifying a particular cis acting nucleic acid element. The larger and more diverse

5 the population, the greater the probability of obtaining  
productive interactions and, therefore, the greater the  
likelihood of obtaining one, or many, cis acting nucleic  
10 acid elements. It is not necessary, however, to use  
5 large diverse populations to practice the methods of the  
invention. For example, populations of isolated nucleic  
acid molecules that are smaller in size or diversity but  
15 which are known or expected to contain cis acting nucleic  
acid elements can similarly be used and result in the  
10 identification of cis acting nucleic acid elements. For  
example, it is possible to identify cis acting nucleic  
20 acid elements from a population as small as two nucleic  
acids. Those skilled in the art will know, or can easily  
determine, the size and diversity of the population of  
15 isolated nucleic acid molecules to be used depending on  
the desired number and types of cis acting nucleic acid  
elements to be identified.

30 A population of at least about  $10^{13}$  different  
nucleic acids that includes all possible molecules of  
20 between 5 and 20 nucleotides in length can readily be  
obtained by synthetic means. For example, by  
35 synthesizing oligonucleotides having each of the four  
naturally-occurring nucleotides at each position, a  
diverse population of approximately  $4^5 + 4^6 + 4^7 + \dots + 4^{20}$  or  
25 approximately  $10^{13}$  different candidate sequences can be  
obtained. Such a population would include virtually  
40 every possible sequence of between 5 and 20 nucleotides  
in length, including virtually every possible cis acting  
nucleic acid element of between 5 and 20 nucleotides in  
45 30 length.

Longer nucleic acid sequences can also be directly  
synthesized, or can be generated by combining shorter  
50 sequences. Methods of combining shorter sequences are

5 known in the art. For example, single-stranded nucleic  
acids with regions of complementarity can be allowed to  
anneal under annealing conditions known in the art. A  
10 polymerization reaction can then be performed to extend  
5 each strand of the oligonucleotide using the overhanging  
portion of the complementary strand as a template.  
Optionally, the strands can be separated, reannealed, and  
15 extension repeated until a diverse population of the  
desired length is achieved.

20 10 As a further example, multiple short double  
stranded DNA sequences can be combined to form longer  
sequences using enzymatic methods known in the art. If  
desired, restriction enzyme sites can be designed in the  
25 flanking sequences or within the nucleic acids containing  
15 the potential cis acting nucleic acid elements.  
Following restriction digestion, random combinations of  
nucleic acid sequences can be ligated together in a  
ligation reaction. Alternatively, random combinations of  
30 double-stranded nucleic acids with blunt ends can be  
20 ligated together in a ligation reaction.

35 If desired, the isolated nucleic acid molecules  
can be flanked on one or both sides with nucleic acid  
sequences with desired properties. For example, an  
isolated nucleic acid molecule can have a restriction  
40 25 enzyme binding consensus sequence or a sequence  
complementary to a primer for amplification by the  
polymerase chain reaction (PCR) at one or both ends.  
These flanking nucleic acid sequences can be used, for  
45 example, to combine or extend nucleic acids as described  
30 above, to amplify nucleic acids sequences by PCR either  
before or after incubation with nucleic acid binding  
factors, or to identify or isolate nucleic acids that  
50

selectively bind to nucleic acid binding factors or compounds.

A diverse population of nucleic acid binding factors is also provided, and is used to contact the diverse population of isolated nucleic acid molecules. Depending on need, the diverse population of nucleic acid binding factors can vary in size and diversity. The larger and more diverse the population, the greater the probability of obtaining productive interactions and, therefore, the greater the likelihood of obtaining one, or many cis acting nucleic acid elements bound to nucleic acid binding factors. It is not necessary, however, to use large diverse populations to practice the methods of the invention. For example, nucleic acid binding factor populations that are smaller in size or diversity but which are known or expected to contain nucleic acid binding factors can similarly be used. Using a population containing as few as two nucleic acid binding factors in the methods of the invention, it is possible to identify one or more cis acting nucleic acid elements. Those skilled in the art will know, or can easily determine, the size and diversity of the nucleic acid binding factor population to be used depending on the desired number and types of cis acting nucleic acid elements and nucleic acid binding factors to be identified.

Depending on need, such as, for example, the type of cis acting nucleic acid element and nucleic acid binding factor one intends to identify, the population of nucleic acid binding factors can be biased to include, for example, nucleic acid binding factors that normally bind to particular types of cis acting nucleic acid elements, that are normally found in particular cell

types, that respond to particular extracellular stimuli,  
or that are localized to particular chromosomal or  
subchromosomal locations.

A source of nucleic acid binding factors can be,  
for example, a cell or subcellular extract obtained by  
biochemical fractionation procedures known in the art. A  
cytoplasmic extract, for example, can be a source of a  
diverse population of nucleic acid binding factors that  
bind, for example, mRNA including, for example, nucleic  
acid binding factors involved in genetic processes such  
as translation, editing, degradation, and the like. A  
nuclear extract, for example, can be a source of a  
diverse population of nucleic acid binding factors that  
bind, for example, hnRNA and single- and double-stranded  
nuclear DNA including, for example, replication factors,  
transcription factors, splicing factors and boundary  
element binding factors. A mitochondrial extract can be  
a source of a diverse population of nucleic acid binding  
factors that bind, for example, mitochondrial DNA. A  
chloroplast extract can be a source of a diverse  
population of nucleic acid binding factors that bind, for  
example, chloroplast DNA.

A source of nucleic acid binding factors can also  
be nucleic acid binding factors bound to nucleic acids,  
either within a cell or obtained from a cell. For  
example, a source of nucleic acid binding factors can be  
cytoplasmic, mitochondrial or nuclear RNA or DNA. A  
source of nucleic acid binding factors can also be a  
preparation of nucleic acids bound to nucleic acid  
binding factors that is isolated from other cellular  
components. For example, where it is desirable to  
identify cis acting nucleic acid elements involved in a  
particular disease or developmental state, nucleic acid

5 binding factors bound to nucleic acids from a particular  
genomic or chromosomal location known to be involved in  
the disease can be used as a source of binding factors.  
10 Therefore, a diverse population of nucleic acid binding  
5 factors bound to nucleic acids can be, for example, bound  
to chromatin, a chromosome, a chromosome arm, a  
transcriptional domain, a gene family or a gene,  
15 depending on the application of the method. A  
transcriptional domain refers to a loop or segment of DNA  
10 that extrudes from chromomeres and that is bounded by cis  
acting boundary elements. Such a structural domain is  
20 often an actively transcribed region of DNA.

If desired, nucleic acid binding factors can be  
released from a nucleic acid preparation and used to  
25 contact the diverse population of isolated nucleic acid  
15 molecules. Methods of releasing nucleic acid binding  
factors bound to a nucleic acid in a nucleic acid  
preparation can be determined for a particular nucleic  
30 acid preparation by those skilled in the art and include,  
20 for example, varying the salt concentration or pH of the  
solution.

35 Diverse populations of nucleic acid binding  
factors can also be obtained by recombinant  
methodologies. One skilled in the art would be able to  
40 25 determine an appropriate source of nucleic acids to  
express to obtain nucleic acid binding factors for a  
particular application of the method. For example, cDNA  
libraries are available or can be produced by known  
45 methods from genes expressed by any desired tissue or  
30 cell source, or in response to any pathogenic or normal  
stimulus.



5  
10  
15  
Depending on the types of cis acting nucleic acid elements one wishes to identify, nucleic acid binding factors can be obtained as described above from cells from different tissues or at different developmental stages. Nucleic acid binding factors can also be obtained from either normal or diseased cells, or following exposure of cells to external stimuli such as therapeutic drugs.

20  
25  
30  
35  
Once the starting populations of isolated nucleic acid molecules and nucleic acid binding factors have been selected and obtained, the populations are combined under conditions that allow the nucleic acid binding factors to selectively bind to the isolated nucleic acid molecules containing cis acting nucleic acid elements. Binding conditions will vary depending on the type and source of nucleic acid binding factors and the type and source of nucleic acids, but can be readily determined. For example, since the affinity and specificity of interactions between nucleic acid binding factors and cis acting nucleic acid elements are generally dependent on the charge of both molecules, one can vary the salt concentration or pH of a buffer to differentially allow binding interactions of particular affinities.

40  
45  
50  
55  
Conditions that allow binding between nucleic acid sequences and nucleic acid binding factors are also designed to ensure that a sufficient concentration of nucleic acids and nucleic acid binding factors are present for a particular application. For example, in one embodiment of the invention, nucleic acid binding factors bound to nucleic acids in a nucleic acid preparation are contacted with a diverse population of isolated nucleic acids. The nucleic acid binding factors will equilibrate between being bound to the cis acting

5 nucleic acid elements present in the nucleic acid  
preparation, and the cis acting nucleic acid elements  
10 present in the diverse population of isolated nucleic  
acid molecules. The distribution of nucleic acid binding  
5 factors between being bound to cis acting nucleic acid  
elements present in the nucleic acid preparation, and  
being bound to cis acting nucleic acid elements in the  
15 isolated population of nucleic acids will depend, for  
example, on the ratio between the number of copies of the  
20 corresponding cis acting nucleic acid elements present in  
the nucleic acid preparation and the number of copies of  
the corresponding cis acting nucleic acid elements in the  
isolated population. An excess of a particular isolated  
25 cis acting nucleic acid element to a cis acting nucleic  
acid element present in the nucleic acid preparation  
would shift the binding equilibrium toward preferential  
binding to the isolated nucleic acid molecules. For  
example, an excess of about 10 to 1, or about  $10^3$  to 1, or  
30 about  $10^6$  to  $10^{10}$  to 1 of isolated cis acting nucleic acid  
elements to cis acting nucleic acid element present in  
the nucleic acid preparation could be used in the  
invention. However, smaller ratios can also be used  
35 without substantially reducing the selectivity of the  
interaction. The use of smaller ratios, including, for  
25 example, equal amounts or less than an excess of isolated  
cis acting nucleic acid elements compared to those in the  
preparation can be advantageous, for example, when  
40 selectively identifying high affinity interactions  
between the cis acting nucleic acid element and nucleic  
30 acid binding factors.

45 As an example, if a chromatin preparation is  
contacted with a diverse population of isolated nucleic  
acid molecules, the number of isolated nucleic acid  
50 molecules is chosen so as to compete with the chromatin

5 for the chromatin-bound factors to a desired extent for a  
particular application. One skilled in the art could  
10 determine the number of copies of each member of the  
diverse population of isolated nucleic acid molecules  
5 required for a particular application of the method.  
Methods known in the art, such as the polymerase chain  
15 reaction, allow production of as many copies of a  
particular isolated nucleic acid sequence as desired.

After allowing isolated nucleic acid molecules to  
10 contact and bind nucleic acid binding factors, nucleic  
acids that selectively bind to nucleic acid binding  
20 factors are identified. These nucleic acids contain one  
or more cis acting nucleic acid elements. Any method for  
identifying nucleic acids that are selectively bound to  
25 nucleic acid binding factors can be used, including  
methods of physically separating bound and unbound  
nucleic acids, as well as methods of distinguishing  
30 between bound and unbound nucleic acids that do not  
require the physical separation of bound from unbound  
20 nucleic acids.

35 Methods of physically separating nucleic acids  
that are bound to binding factors from nucleic acids that  
are unbound are known in the art. For example, nucleic  
acids that are bound to nucleic acid binding factors and  
40 25 those that are unbound can be separated by virtue of  
size, shape, charge or density of the bound complex as  
compared to unbound nucleic acids. For example, nucleic  
acids bound to nucleic acid binding factors will pass  
45 through a chromatography column at a different rate than  
30 unbound nucleic acids. Appropriate chromatography resins  
can be determined by those skilled in the art for a  
particular application. Additionally, depending on the  
50 nature of the nucleic acid binding factor, a nucleic acid

bound to a nucleic acid binding factor can have a greater or lesser density than an unbound nucleic acid, and can be separated from unbound nucleic acids by known methods of density centrifugation. Furthermore, bound and unbound nucleic acids will have different electrophoretic mobilities, and can be separated by methods known in the art such as electrophoretic mobility shift assays (EMSA). If desired, the bound nucleic acids can be isolated, stored, amplified, sequenced or used as described below.

Furthermore, it is known that a nitrocellulose membrane will selectively retain double-stranded DNA bound to proteinaceous nucleic acid binding factors, but will allow unbound DNA to pass through the filter. Therefore, following binding of isolated nucleic acid molecules with nucleic acid binding factors, the binding reaction can be filtered through a nitrocellulose filter. DNAs that are bound to nucleic acid binding factors are retained on a nitrocellulose filter. These DNAs contain cis acting nucleic acid elements. If desired, the retained nucleic acids can be eluted from the nitrocellulose membrane and stored, amplified, sequenced or used as described below. One skilled in the art can also vary buffer conditions to selectively retain single-stranded nucleic acid sequences bound to nucleic acid binding factors on nitrocellulose filters while allowing unbound nucleic acid sequences to pass through the filters. The retained nucleic acids contain cis acting nucleic acid elements. One skilled in the art could also modify such as assay by, for example, varying the type of membrane, to selectively retain nucleic acids bound to non-proteinaceous nucleic acid binding factors.

Methods of distinguishing between nucleic acids that are bound to nucleic acid binding factors and those

5 that are unbound, which do not require the physical  
separation of bound from unbound nucleic acids, are  
10 similarly known in the art. A method of distinguishing  
between bound nucleic acids and unbound nucleic acids  
5 takes advantage of properties that distinguish bound  
nucleic acids as compared to unbound nucleic acids such  
as, for example, nuclease resistance. As one example of  
15 the use of nuclease resistance to distinguish bound from  
unbound nucleic acids, a diverse population of isolated  
10 double-stranded DNA can be flanked, at one or both ends,  
with a sequence containing the binding site of a  
20 restriction enzyme that is known, or can be designed, to  
cut at a site at a distance away from the binding site.  
Both ends of the nucleic acid also contain sequences that  
25 are complementary to PCR primers. Following binding  
between isolated nucleic acid molecules and nucleic acid  
binding factors, the reaction mixture is further  
incubated with such a restriction enzyme under conditions  
30 that allow cleavage of DNA at the restriction enzyme  
cleavage site only if the cleavage site is not bound to a  
nucleic acid binding factor. Thus, DNA that is unbound  
is cleaved, and bound DNA is not cleaved. Uncleaved DNA  
35 therefore retains PCR primer sites at both ends of the  
cis acting nucleic acid element and can be amplified by  
25 PCR, whereas cleaved DNA only has a single primer site  
and can not be amplified by PCR. If desired, the nucleic  
40 acid binding factor and restriction enzyme can be removed  
by methods known in the art, such as by appropriately  
varying the buffer conditions. A PCR reaction is then  
30 performed, which amplifies only those nucleic acids that  
45 were bound to nucleic acid binding factors. These  
nucleic acids contain cis acting nucleic acid elements.

50 Restriction enzymes that cleave at a distance of  
about 5 to about 30 nucleotides away from the binding

5 site are commercially available. Such enzymes include,  
for example, BbvI, BcgI, BciVI, BpmI, BseRI, BsmFI, FokI,  
10 HgaI, HphI, MboII, MnlI and SfaNI, each of which is  
available from New England BioLabs, Inc. Using knowledge  
5 of restriction enzyme structure, it is also possible to  
design restriction enzymes that combine a desired binding  
site specificity with a desired cleavage site specificity  
15 and cleavage site distance.

For certain methods of distinguishing between  
20 bound and unbound nucleic acids, it may be desirable to  
detectably label either the diverse population of nucleic  
acids or the diverse population of nucleic acid binding  
factors. Detectable labels include moieties such as, for  
25 example, enzymes, radioisotopes, fluorochromes,  
chemiluminescent markers, and biotin, which can be  
15 incorporated into isolated nucleic acid molecules and  
nucleic acid binding factors, or incorporated by  
metabolic labeling into nucleic acids and nucleic acid  
30 binding factors *in vivo* or in cultured cells. A  
detectable label can also be a tag that can be  
20 specifically recognized by a binding moiety, such as, for  
example, an antibody.  
35

For certain applications of the method, such as  
25 high-throughput screening for therapeutic compounds and  
for diagnostic procedures, it is advantageous to provide  
40 the diverse population of nucleic acids on a solid  
support. The diverse population of nucleic acids can be  
synthesized on, or subsequently attached to, solid  
45 supports such as beads, pins, resins or chips. Nucleic  
acids attached to solid supports can be contacted with  
nucleic acid binding factors; those nucleic acid binding  
factors that are not specifically bound to nucleic acids  
50 are removed, and the nucleic acids, both bound and  
55

unbound, remain attached to the solid support. The bound nucleic acids can be detected, for example, by virtue of the detectable label present in either the nucleic acid or the nucleic acid binding factor, or by virtue of another inherent detectable property, such as charge, size or nuclease resistance, that distinguishes bound from unbound nucleic acids.

For example, the fluorescence of a fluorescently labeled nucleic acid can be quenched by binding to a nucleic acid binding factor, and this quenching can be detected. Similarly, the amount of chemiluminescent signal or radioactivity of a nucleic acid that can be detected can be altered by binding to a nucleic acid binding factor. Additionally, binding of nucleic acid binding factor can protect a nucleic acid from degradation by nucleases, and the undegraded nucleic acids can be detected by virtue of their detectable labels.

It is not necessary to be able to directly isolate a nucleic acid that is bound to a nucleic acid binding factor in order to identify it, if the corresponding sequence of the nucleic acid that was bound to the binding factor is known. For example, nucleic acids can be synthesized on solid supports in arrays, with nucleic acids of known sequences present at known locations. Therefore, any property that identifies selectively bound nucleic acids from unbound nucleic acids in a diverse population of nucleic acids present in an array of nucleic acids can be used to identify cis acting nucleic acid elements. Nucleic acid chips and automated detection procedures are particularly advantageous in high-throughput screening procedures for identifying cis acting nucleic acid elements, nucleic acid binding

factors, and compounds that bind cis acting nucleic acid elements and nucleic acid binding factors.

Solid phase oligonucleotide synthesis methods are known in the art (see, for example, J. Weiler et al., *Anal. Biochem.* 243:218 (1996) and U. Maskos et al., *Nucleic Acids Res.* 20(7):1679 (1992); T. Atkinson et al., *Solid-Phase Synthesis of Oligodeoxyribonucleotides by the Phosphitetriester Method*, in *Oligonucleotide Synthesis* 35 (M.J. Gait ed., 1984), as are methods for synthesizing arrays of oligonucleotides (see, for example, United States Patent No. 5,474,796; International Publication No. WO 95/25116; Blanchard et al., "High-density oligonucleotide arrays" *Biosensors & Bioelectronics* 11(6/7):687-690 (1996)).

The above methods of distinguishing between nucleic acids that are bound to nucleic acid binding factors and those that are not can be used individually, or in any combination or order, to identify nucleic acids containing cis acting nucleic acid elements.

Once the sequences of one or a plurality of isolated nucleic acid molecules containing cis acting nucleic acid elements is determined, any desired set or subset thereof can be synthesized, using methods known in the art, and used in a variety of therapeutic, diagnostic and screening methods. The cis acting nucleic acid elements within the isolated nucleic acid molecules can be determined, if desired, by means known in the art. For example, known methods of nucleic acid "footprinting" can be used. A nucleic acid can be detectably labeled and contacted with a nucleic acid binding factor or population of nucleic acid binding factors. The nucleic acid is then partially digested with a nuclease. The



5 sequences that are protected from nuclease digestion by  
the bound nucleic acid binding factor are the cis acting  
10 nucleic acid elements.

15 If desired, the sequences of isolated cis acting  
5 nucleic acid elements identified by a method of the  
invention can be directly compared with cis acting  
15 nucleic acid elements found in cellular or viral DNA or  
RNA. Such comparison is advantageous, for example, in  
determining the extent to which a cis acting nucleic acid  
20 element identified by a method of the invention is  
identical to a cis acting nucleic acid element found in  
naturally occurring populations of nucleic acids. Such  
comparison also advantageously allow the determination of  
25 which nucleic acids are regulated by particular cis  
15 acting nucleic acid elements. These regulated nucleic  
acids can include previously unknown or uncharacterized  
genes involved in disease or development, which can  
30 themselves be used in therapeutic and diagnostic  
procedures.

35 20 Several methods are known in the art that can be  
used to compare sequences of isolated cis acting nucleic  
acid elements to cis acting nucleic acid elements found  
in cellular or viral DNA or RNA. For example, the  
partial or complete genomic sequences of a variety of  
40 25 different organisms, including humans, are available in  
databases. These databases can be searched for identical  
or substantially similar sequences to the cis acting  
nucleic acid elements identified by a method of the  
45 invention. The regulated genes can be identified and  
30 produced by recombinant or synthetic means known in the  
art.

50

55

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

Additionally, populations of nucleic acids cloned in, for example, phage, plasmid, cosmid or YAC libraries are available or can be prepared by methods known in the art. These libraries can be screened using methods known in the art, such as nucleic acid hybridization, to determine the cis acting nucleic acid elements and flanking sequences in the cellular or viral nucleic acids that are substantially similar to the cis acting nucleic acid elements identified by the methods of the invention.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55

Furthermore, the location of one or a plurality of cis acting nucleic acid elements within a particular cell compartment or within a particular chromosome can be advantageously used to characterize the cis acting nucleic acid elements and the nucleic acids they regulate. For example, depending on the starting population of isolated nucleic acid molecules and nucleic acid binding factors, several types of cis acting nucleic acid elements could be simultaneously identified. Therefore, by examining the location of hybridization of a cis acting nucleic acid element to the cellular nucleic acids, the type of cis acting nucleic acid element and the location of the regulated nucleic acids can be determined. For example, boundary elements, elements that bind telomeres and elements that bind transcription factors could be distinguished by knowing where each element mapped to the chromosomes. Similarly, RNA elements that are present in mRNA as compared to hnRNA could be distinguished by virtue of their intracellular location. Such methods of mapping nucleic acid sequences to particular nucleic acid locations are known in the art and include, for example, fluorescence *in situ* hybridization (FISH).

5 The methods of the invention for identifying and  
isolating cis acting nucleic acid elements that are bound  
to nucleic acid binding factors also simultaneously  
10 provide for the identification and isolation of nucleic  
5 acid binding factors that selectively bind cis acting  
nucleic acid elements. Therefore, the invention provides  
a method of isolating a nucleic acid binding factor. The  
15 method involves contacting a diverse population of  
nucleic acid binding factors with a diverse population of  
10 isolated nucleic acid molecules under conditions that  
allow nucleic acid binding factors to selectively bind  
20 nucleic acids, and isolating one or more nucleic acid  
binding factors that selectively bind one or more  
isolated nucleic acid molecules. The source and  
25 diversity of the populations of nucleic acid binding  
factors and isolated nucleic acid molecules can be  
determined by those skilled in the art, as described  
previously, based on the type and number of nucleic acid  
30 binding factors that it is desired to isolate in a  
20 particular application of the method.

Following contacting the populations of isolated  
35 nucleic acid molecules and nucleic acid binding factors,  
the isolated nucleic acid molecules that are selectively  
bound by nucleic acid binding factors are separated from  
25 unbound nucleic acids. As described previously, methods  
are known in the art to physically separate nucleic acids  
40 that are bound to nucleic acid binding factors from  
nucleic acids that are unbound. Such methods include,  
for example, filtration, chromatography, electrophoresis  
45 and centrifugation. The selectively bound nucleic acid  
30 binding factors are dissociated from the nucleic acids  
they bind and are isolated. Methods of dissociating  
nucleic acid binding factors from nucleic acids are known  
50

in the art and include, for example, varying the salt or detergent concentration or the pH of the buffer.

Once isolated, the nucleic acid binding factor of interest can be produced in large quantity from a diverse population of nucleic acid binding factors using, for example, its corresponding cis acting nucleic acid element or other binding agent, such as a specific antibody, as an affinity reagent. Furthermore, if a nucleic acid binding factor is a protein, the sequence of the encoding gene can be readily determined and the nucleic acid binding factor can be recombinantly produced.

The site of interaction between a nucleic acid binding factor and other binding factors in a binding complex, and the site of interaction between a nucleic acid binding factor and its corresponding cis acting nucleic acid element, also can be determined using methods known in the art. Knowledge about these sites of interaction can be used to design therapeutic compounds that alter or disrupt these interactions.

The genetic circuitry of cells and viruses controls cell and organismal behavior, including, for example, proliferation, differentiation and pathogenicity. Therefore, being able to modulate the control properties, dynamics or behavior of the genetic circuitry or to modify the genetic circuitry directly, of a host cell or a pathogen in a controlled way, in order to alter nucleic acids that mediate these processes, can be advantageous for therapy. For example, modulating the control properties, dynamics or behavior of the genetic circuitry of a cell, or modifying the genetic circuitry directly, can be used to modulate the proliferation,

5 differentiation, susceptibility to disease or  
susceptibility to drugs of the cell, depending on the  
particular therapeutic application. Modulating the  
10 control properties, dynamics or behavior of the genetic  
5 circuitry of a pathogen, or modifying its genetic  
circuitry directly, can also be used to modulate the  
infectivity, pathogenicity or drug resistance of the  
15 pathogen.

The identification of cis acting nucleic acid  
10 elements and nucleic acid binding factors provides a  
means of rapidly identifying compounds that can alter the  
control properties, dynamics or behavior of the genetic  
circuitry of a cell or virus for therapeutic purposes.  
25 The identification of cis acting nucleic acid elements  
15 that modulate a genetic activity of nucleic acids  
involved in a pathological condition also provides a  
means of inserting, removing or replacing the cis acting  
nucleic acid elements to directly modify the genetic  
30 circuitry of a cell for therapeutic purposes.

20 The methods of the invention provide for the  
identification of therapeutic compounds that can target  
any nucleic acid or group of nucleic acids of interest  
that contain one or more cis acting nucleic acid  
elements. Such therapeutic compounds include, for  
40 25 example, analogs of cis acting nucleic acid elements,  
analogous of nucleic acid binding factors, compounds that  
bind to either cis acting nucleic acid elements or  
nucleic acid binding factors or both, as well as cis  
45 acting nucleic acids and nucleic acid binding factors  
30 themselves. These therapeutic compounds can, for  
example, compete with an endogenous cis acting nucleic  
acid element for binding to a nucleic acid binding  
50 factor, or compete with a nucleic acid binding factor for

5 binding with its corresponding cis acting nucleic acid  
element. These compounds can also physically disrupt the  
10 binding of an endogenous cis acting nucleic acid element  
to its corresponding nucleic acid binding factor or  
5 disrupt the binding between two or more nucleic acid  
binding factors.

15 Altering the regulation of nucleic acids  
associated with disease can prevent or treat disease.  
Compounds that target cis acting nucleic acid elements  
10 and nucleic acid binding factors involved in particular  
20 diseases can be identified and used to enhance, inhibit,  
alter, antagonize or mimic the regulation of a nucleic  
acid known or predicted to be associated with disease.  
For example, cis acting nucleic acid elements or nucleic  
25 acid binding factors that are known or expected to  
modulate one or a plurality of nucleic acids involved in  
cancer, degenerative diseases, genetic disorders, immune  
30 disorders, bacterial and viral infectious diseases and  
the like, can be used in the methods described below to  
20 identify specific therapeutic compounds that will target  
the corresponding regulated nucleic acid. These  
35 therapeutic compounds can beneficially alter a genetic  
activity of the nucleic acid, such as, for example, its  
structural integrity, transcription, translation, or  
25 replication, so as to ameliorate or prevent the disease.

40 The isolated nucleic acid molecules or the nucleic  
acid binding factors, or both, in the exemplary methods  
of identifying therapeutic compounds described below, can  
45 be biased populations that include cis acting nucleic  
30 acid elements or nucleic acid binding factors that are  
known or predicted to regulate nucleic acids involved in  
a disease. The compounds so obtained would be expected  
50 to preferentially include compounds that are selective  
for the nucleic acids involved in the particular disease.

5 Alternatively, the starting populations can be large,  
random populations of nucleic acids and nucleic acid  
10 binding factors. In the latter case, it would be  
expected that a library of compounds would be obtained,  
5 only a few of which would be selective for any particular  
nucleic acid or nucleic acid binding factor. However,  
15 the library of compounds obtained using the methods of  
the invention can readily be screened to determine which  
subset of compounds alters the regulation of any nucleic  
20 acid of interest.

20 Methods of screening to determine that a compound  
alters the regulation of a particular nucleic acid can be  
determined by those skilled in the art depending on the  
25 nucleic acid and its properties. For example, the  
15 affinity and selectivity of a compound for binding to a  
particular cis acting nucleic acid element or nucleic  
acid binding factor could be determined using a binding  
30 competition assay. Likewise, the effect of a compound on  
the regulation of a nucleic acid could be determined by  
20 examining the expression of the mRNA or protein encoded  
by the regulated nucleic acid. Furthermore, the effect  
35 of the compound on a property of a cell, such as growth,  
differentiation or apoptosis, that depends on the  
expression of the gene, could be determined.

40 25 Compounds that selectively bind to nucleic acid  
binding factors, such that they can be selectively  
displaced by isolated nucleic acid molecules, are analogs  
of cis acting nucleic acid elements. Such compounds are  
45 potential therapeutic agents that can alter a genetic  
30 activity modulated by a cis acting nucleic acid element  
of which the compound is an analog. Therefore, the  
invention provides a method of identifying a cis acting  
50 nucleic acid element analog. The method involves

5 contacting a diverse population of nucleic acid binding  
factors with a diverse population of compounds under  
10 conditions that allow the compounds to selectively bind  
the nucleic acid binding factors. One or more of the  
5 nucleic acid binding factors selectively bound to one or  
more of the compounds is contacted with one or more  
isolated nucleic acid molecules under conditions that  
15 allow one or more of the isolated nucleic acid molecules  
to selectively displace one or more of the selectively  
20 bound compounds. The isolated nucleic acid molecules or  
the nucleic acid binding factors, or both, can correspond  
to or regulate nucleic acids that are known or expected  
to play a role in a disease of interest. The displaced  
compounds are identified and characterized as *cis* acting  
25 nucleic acid element analogs. Such a method further  
provides for the identification of one or more of the  
isolated nucleic acid molecules that selectively  
displaces one or more of the selectively bound compounds.  
30 An isolated nucleic acid molecule that selectively  
20 displaces one or more of the selectively bound compounds  
is characterized as a nucleic acid containing a *cis*  
acting nucleic acid element.

35 Compounds that selectively bind to isolated  
nucleic acid molecules or to nucleic acid binding factors  
25 in a nucleic acid binding factor complex, such that they  
can be displaced by selectively binding to nucleic acid  
40 binding factors, are analogs of nucleic acid binding  
factors. Such compounds are potential therapeutic agents  
that can alter a genetic activity modulated by a *cis*  
45 30 acting nucleic acid element that binds a nucleic acid  
binding factor of which the compound is an analog.  
Therefore, the invention also provides a method of  
identifying nucleic acid binding factor analogs. In one  
50 embodiment, the method consists of contacting a diverse



5 population of compounds with a diverse population of  
isolated nucleic acid molecules under conditions that  
allow the compounds to selectively bind the isolated  
10 nucleic acid molecules. One or more of the isolated  
5 nucleic acid molecules selectively bound to one or more  
of the compounds is contacted with one or more nucleic  
acid binding factors under conditions that selectively  
15 displace one or more of the selectively bound compounds  
from one or more of the bound nucleic acids. The  
10 isolated nucleic acid molecules or the nucleic acid  
binding factors, or both, can correspond to or regulate  
20 nucleic acids that are known or expected to play a role  
in a disease of interest. The displaced compounds are  
identified, and are characterized as nucleic acid binding  
25 factor analogs. The method further provides for the  
identification of one or more nucleic acid binding  
factors that displaces one or more of the selectively  
bound compounds.

30 In a further embodiment of the above method,  
20 compounds that selectively bind either to cis acting  
nucleic acid elements or to nucleic acid binding factors  
35 in a nucleic acid binding factor complex or to both can  
be simultaneously identified. The method involves  
contacting a diverse population of compounds with a  
25 diverse population of isolated nucleic acid molecules  
bound to nucleic acid binding factors under conditions  
40 that allow the compounds to selectively bind to either  
the isolated nucleic acid molecules or to the nucleic  
acid binding factors. One or more of the isolated  
45 30 nucleic acid molecules selectively bound to nucleic acid  
binding factors and selectively bound to one or more  
compounds is contacted with one or more nucleic acid  
binding factors under conditions that allow one or more  
50 of the nucleic acid binding factors to selectively

5                   displace one or more of the selectively bound compounds.  
The isolated nucleic acid molecules or the nucleic acid  
10                   binding factors, or both, can correspond to or regulate  
nucleic acids that are known or expected to play a role  
5                   in a disease of interest. The displaced compounds are  
identified, and are characterized as nucleic acid binding  
15                   factor analogs. The displaced compounds can further be  
characterized to determine whether they bind to a cis  
acting nucleic acid element or to a nucleic acid binding  
20                   factor in a complex of nucleic acid binding factors.

20                   Compounds that selectively bind to cis acting  
nucleic acid elements can also be used as therapeutic  
agents to alter the activity of nucleic acids modulated  
25                   by cis acting nucleic acid elements. Therefore, the  
invention also provides a method of identifying compounds  
that bind cis acting nucleic acid elements. The method  
involves contacting a plurality of isolated nucleic acid  
30                   molecules, wherein each nucleic acid comprises one or  
more cis acting nucleic acid elements, with a diverse  
population of compounds under conditions that allow the  
compounds to selectively bind the isolated nucleic acid  
35                   molecules. The compounds that selectively bind one or  
more isolated nucleic acid molecules containing one or  
25                   more cis acting nucleic acid elements are identified.

40                   As described previously, the isolated nucleic acid  
molecules containing cis acting nucleic acid elements can  
correspond to nucleic acids that are known or expected to  
play a role in a disease of interest, or can be a large,  
45                   30 random population. A compound identified by the method  
can be tested for its ability to bind a cis acting  
nucleic acid element of interest by direct or indirect  
assays known in the art. Such assays include, for  
50

example, binding assays, reporter assays, and functional assays that measure the effect of introduction of the compound on a property of the cell.

The invention also provides a method of identifying compounds that selectively displace binding of a cis acting nucleic acid element to a nucleic acid binding factor or of a nucleic acid binding factor to another nucleic acid binding factor. The method involves contacting a plurality of isolated nucleic acid molecules selectively bound to nucleic acid binding factors, with a diverse population of compounds under conditions that allow the compounds to selectively displace one or more of the selectively bound nucleic acid binding factors from one or more of the bound nucleic acids or from one or more of the bound nucleic acid binding factors in the binding factor complex. The isolated nucleic acid molecules containing cis acting nucleic acid elements or the nucleic acid binding factors, or both, can correspond to or regulate nucleic acids that are known or expected to play a role in a disease of interest. The isolated nucleic acid molecules can be selected to each contain one or more cis acting nucleic acid elements. The compounds that selectively displace one or more of the bound nucleic acid binding factors from one or more of the bound nucleic acids or from one or more of the bound nucleic acid binding factors in the binding factor complex are identified. Such a compound can, for example, bind to the site of interaction between the cis acting nucleic acid element and the nucleic acid binding factor and be, therefore, either a cis acting nucleic acid element analog or a nucleic acid binding factor analog. Such a compound can also, for example, bind to the site of interaction between two or more nucleic acid binding factors within a nucleic acid binding factor

5 complex. Alternatively, such a compound can bind  
elsewhere on the cis acting nucleic acid element or  
elsewhere on one or more of the nucleic acid binding  
10 factors, so long as binding between a nucleic acid  
5 binding factor and either a cis acting nucleic acid  
element or another nucleic acid binding factor is  
selectively modified or displaced by binding of the  
15 compound.

The methods of the invention described above can  
20 be used to identify compounds that are selective for many  
different nucleic acids as well as compounds that target  
only a very limited number of nucleic acids. As  
described previously, some of the cis acting nucleic acid  
25 elements that regulate a particular nucleic acid will  
15 likely also be involved in the regulation of numerous  
other nucleic acids. Therefore, a therapeutic compound  
that binds to that cis acting nucleic acid element or its  
30 corresponding nucleic acid binding factor may have an  
effect on the regulation of many nucleic acids other than  
20 the intended target nucleic acid. However, a particular  
combination of cis acting nucleic acid elements will be  
35 relatively specific for a particular nucleic acid or  
family of nucleic acids. Therefore, the invention also  
provides for the identification of therapeutic agents  
25 that are specific for one or several nucleic acids by  
40 using isolated nucleic acid molecules that include a  
combination of cis acting nucleic acid elements in the  
methods described above. The cis acting nucleic acid  
45 elements in the combination of cis acting nucleic acid  
30 elements can be linked by the naturally occurring  
intervening sequences. Alternatively, so as to provide  
for a convenient overall nucleic acid length, non-native  
50 intervening sequences can be introduced between the cis  
acting nucleic acid elements. Using the methods

5 described above, therapeutic compounds that selectively  
bind to the combination of cis acting nucleic acid  
10 elements, or compounds that selectively bind to or  
displace the combination of nucleic acid binding factors,  
5 can be identified.

15 The above methods of identifying compounds that  
can be used as therapeutic agents take advantage of the  
ability to distinguish between nucleic acids that are  
10 selectively bound to particular compounds or binding  
factors, and nucleic acids that are either unbound or  
20 bound to different compounds or binding factors. Any  
method of distinguishing bound from unbound nucleic acids  
can be used in the above methods of identifying  
25 therapeutic compounds that bind cis acting nucleic acid  
elements and nucleic acid binding factors, such as those  
described previously. Such methods can be automated by,  
for example, providing arrays of isolated nucleic acid  
30 molecules on solid supports. Similarly, arrays of  
compounds on solid supports can be provided. The  
compounds, the nucleic acid binding factors, or the  
nucleic acids can be detectably labeled by methods known  
35 in the art. Additionally, isolated nucleic acid  
molecules that are bound to particular compounds can  
25 differ from unbound nucleic acids or nucleic acids bound  
to different compounds or nucleic acid binding factors in  
their ability to be retained on filters such as  
40 nitrocellulose filters, and can differ in charge, size,  
density, electrophoretic mobility and resistance to  
30 nucleases.

45 Compounds, nucleic acid binding factors, and  
isolated nucleic acids can be removed from the molecules  
they selectively bind for further characterization, if  
50 desired. Alternatively, pools of such molecules can be

repeatedly subdivided until one or a plurality of selectively bound or selectively displaced molecules is isolated or identified.

The invention also provides a plurality of isolated nucleic acid molecules, wherein each isolated nucleic acid molecule contains one or more cis acting nucleic acid elements. Such a plurality of isolated nucleic acid molecules containing cis acting nucleic acid elements can contain, for example, between about 2-5 different isolated nucleic acid molecules, or between about 6-10 different isolated nucleic acid molecules. The plurality of isolated nucleic acids can also contain between about 11-20 different isolated nucleic acid molecules or greater than about 20 different isolated nucleic acid molecules. The number of isolated nucleic acid molecules will depend on the type of nucleic acids in the plurality and the intended use of the plurality. These nucleic acids can be attached to a solid support, if desired, and advantageously used for automated screening and diagnostic procedures.

A plurality of isolated nucleic acid molecules containing cis acting nucleic acid elements can be identified and obtained, for example, by the methods described above. The plurality can be produced in abundance by, for example, chemical synthesis or by amplification by the polymerase chain reaction. If desired, isolated cis acting nucleic acid elements can be synthesized with various amounts of adjacent sequences. These adjacent sequences can be used, for example, in the detection, amplification, cloning or further modification of the sequences. As described above, a plurality of isolated nucleic acid molecules containing cis acting nucleic acid elements can be, for example, a set of

5 isolated transcription factor binding elements, such as  
enhancers and promoters; a set of isolated replication  
factor binding elements, such as origins of replication;  
10 a set of isolated restriction or modification enzyme  
binding sites; or any other set of nucleic acid cis  
5 acting elements that regulates a desired genetic activity  
of nucleic acids.

15 As described above, a plurality of isolated  
nucleic acid molecules containing cis acting nucleic acid  
elements can be characteristic of, for example, a  
20 particular cell type, a particular disease or  
developmental state of a cell, or a particular response  
to external stimuli. A plurality of nucleic acids  
containing cis acting nucleic acid elements can also be  
25 characteristic of a particular subset of cellular nucleic  
acids, such as a chromosomal region that maps to a  
disease locus.

30 The invention also provides a plurality of  
isolated nucleic acid molecules bound to nucleic acid  
binding factors, wherein each isolated nucleic acid  
20 molecule contains one or more cis acting nucleic acid  
elements. Such a plurality of isolated nucleic acid  
35 molecules bound to nucleic acid binding factors can  
contain, for example, between about 2-5 different  
isolated nucleic acid molecules, or between about 6-10  
40 different isolated nucleic acid molecules. The plurality  
of isolated nucleic acids can also contain between about  
11-20 different isolated nucleic acid molecules or  
45 greater than about 20 different isolated nucleic acid  
molecules. The number of isolated nucleic acid molecules  
30 bound to nucleic acid binding factors will depend on the  
type of nucleic acids and nucleic acid binding factors in  
50 the plurality and the intended use of the plurality.

5 These nucleic acids or nucleic acid binding factors can  
be attached to a solid support, if desired, and  
10 advantageously used for automated screening and  
diagnostic procedures. As described above, such a  
5 plurality can be used, for example, to identify  
therapeutic compounds that can selectively modify or  
displace the binding of a cis acting nucleic acid element  
15 to a nucleic acid binding factor or that can selectively  
modify or displace the binding between two or more  
10 nucleic acid binding factors.

20 The invention also provides a plurality of  
isolated nucleic acid binding factors that includes at  
least about 15 different isolated nucleic acid binding  
25 factors. The plurality of isolated nucleic acid binding  
factors can also contain between about 16-25 different  
15 isolated nucleic acid binding factors, preferably between  
about 26-50 different isolated nucleic acid binding  
factors, and more preferably greater than about 51  
30 different isolated nucleic acid binding factors. The  
20 number of isolated nucleic acid binding factors in the  
plurality will depend on the type of nucleic acid binding  
factors in the plurality and the intended use of the  
35 plurality. If desired, the plurality of isolated nucleic  
acid binding factors can be attached to a solid support,  
25 and advantageously used for automated screening and  
40 diagnostic procedures.

45 The invention also provides a plurality of cis  
acting nucleic acid analogs. Such a plurality of cis  
acting nucleic acid analogs can include between about 2-5  
30 different isolated cis acting nucleic acid element  
analog, or between about 6-10 different isolated cis  
acting nucleic acid element analogs. The plurality of  
50 cis acting nucleic acid analogs can also contain ,



5 between about 11-20 different isolated cis acting nucleic  
acid element analogs or greater than about 20 different  
10 isolated cis acting nucleic acid element analogs. These  
analogues can be compounds obtained, for example, by the  
5 methods of the invention and are potential therapeutic  
agents that can be used to alter the interactions between  
the cis acting nucleic acid elements they mimic and  
15 nucleic acid binding factors.

The invention further provides a plurality of  
20 nucleic acid binding factor analogs. Such a plurality of  
cis acting nucleic acid analogs can include between about  
2-5 different isolated cis acting nucleic acid element  
analogues, or between about 6-10 different isolated cis  
25 acting nucleic acid element analogs. The plurality of  
isolated cis acting nucleic acid element analogs can also  
15 contain, between about 11-20 different isolated cis  
acting nucleic acid element analogs or greater than about  
30 20 isolated cis acting nucleic acid element analogs.  
These analogs can be compounds obtained, for example, by  
20 the methods of the invention and are potential  
therapeutic agents that can be used to alter the  
35 interactions between the nucleic acid binding factors  
they mimic and either cis acting nucleic acid elements or  
other nucleic acid binding factors within a complex of  
25 nucleic acid binding factors.

40 As described previously, the invention provides  
for the identification of cis acting nucleic acid  
elements and nucleic acid binding factors that regulate  
45 or modulate the genetic activity of nucleic acids that  
cause or are involved pathological conditions. The  
30 methods of the invention also provide for the  
identification of therapeutic compounds, including cis  
50 acting nucleic acid elements, nucleic acid binding

5 factors and their analogs, that can be used  
therapeutically to alter the genetic activity of these  
10 nucleic acids involved in pathological conditions.  
Therefore, the invention provides a method of treating a  
5 pathological condition in an individual. The method  
involves administering to an affected individual an  
effective amount of one or more therapeutic agents that  
15 selectively alter the ability of one or more cis acting  
nucleic acid elements to regulate a genetic activity of  
10 one or more nucleic acids involved in the pathological  
condition.

20  
A pathological condition mediated by the  
dysregulation of one or more nucleic acids can be treated  
25 by a method of the invention. For example, a therapeutic  
15 compound can be administered to either selectively  
increase or selectively decrease a genetic activity of  
one or more nucleic acids that is dysregulated in the  
30 cells of the diseased individual, as required.  
Similarly, a pathological condition mediated by a virus  
20 or bacteria can be treated by administering a compound  
that selectively alters a genetic activity of the  
35 pathogen.

The nucleic acids involved in the pathological  
condition are known in the art or are determined, for  
40 25 example, as described below using the knowledge that cis  
acting nucleic acid elements are present in the vicinity  
of actively transcribed genes. The appropriate genetic  
activity to target using a method of the invention can be  
45 determined by those skilled in the art and will depend on  
30 the underlying disease mechanism for a particular  
disease. As one example, cancer can be treated by  
administering a therapeutic compound of the invention  
50 that selectively targets oncogene transcription. As a

5 further example, a viral infection can be treated by  
administering a compound of the invention that  
10 selectively targets viral replication.

A therapeutic agent can be formulated into a  
5 pharmaceutical composition that is convenient for  
delivering the agent to the target cells and to the  
15 target location within the cell, such as, for example,  
the nucleus or cytoplasm. Such pharmaceutical  
compositions contain the therapeutic agent together with  
20 a pharmaceutically acceptable carrier. Pharmaceutically  
acceptable carriers are well known in the art and include  
aqueous solutions such as water, physiologically buffered  
saline or other solvents or vehicles such as glycols,  
25 glycerol, oils such as olive oil or injectable organic  
15 esters and liposomes.

A pharmaceutically acceptable carrier can contain  
30 physiologically acceptable compounds that act, for  
example, to stabilize or increase the absorption of the  
therapeutic agent. Such physiologically acceptable  
20 compounds include, for example, carbohydrates, such as  
35 glucose, sucrose or dextrans, antioxidants, such as  
ascorbic acid or glutathione, chelating agents, low  
molecular weight proteins or other stabilizers or  
excipients. One skilled in the art would know that the  
40 25 choice of a pharmaceutically acceptable carrier,  
including a physiologically acceptable compound, depends,  
for example, on the nature of the therapeutic agent and  
on the route of administration.

The therapeutic agent also can be incorporated, if  
30 desired, into liposomes, which consist of phospholipids  
or other lipids, and are nontoxic, physiologically  
50 acceptable and metabolizable carriers that are relatively  
simple to make and administer. Targeting of a

5 therapeutic agent encapsulated in liposomes to a cell or  
tissue in an individual can be passive or active.  
Passive targeting, for example, utilizes the tendency of  
10 liposomes to accumulate in cells of the  
5 reticuloendothelial system (RES) and in organs such as  
the liver, which contain sinusoidal capillaries. Active  
targeting, in comparison, involves alteration of the  
15 liposome by coupling a specific ligand such as a  
monoclonal antibody, a sugar, a glycolipid or a protein  
10 such as a ligand for a receptor expressed by the target  
cells.

20 A nucleic acid therapeutic agent, or an encoded  
polypeptide, can be contained in a vector known in the  
art, such as a plasmid, cosmid, or viral vector. Viral  
25 vectors such as retroviral vectors, adenovirus vectors,  
herpes simplex virus vectors, vaccinia virus and the like  
are particularly useful for the administration of nucleic  
acid therapeutic agents and encoded polypeptides. The  
30 choice of vector and route of administering the vector  
will depend, for example, on the particular target cells,  
20 and can be determined by those skilled in the art.

35 A therapeutic agent that modulates genetic  
activities mediated by cis acting nucleic acid elements  
can be administered to an individual by various routes  
40 including, for example, orally or parenterally, such as  
intravenously, intramuscularly, subcutaneously,  
intraorbitally, intracapsularly, intraperitoneally,  
intrarectally intracisternally or by passive or  
45 facilitated absorption through the skin using, for  
30 example, a skin patch or transdermal iontophoresis,  
respectively. Furthermore, a therapeutic agent can be  
administered by injection, intubation, orally or  
50 topically, the latter of which can be passive, for

example, by direct application of an ointment or powder, or active, for example, using a nasal spray or inhalant.

Compounds identified as described above as therapeutic agents can be further modified using known methods so as to have, for example, enhanced stability or bioavailability, or to have optimal affinity for a cis acting nucleic acid element or a nucleic acid binding factor. A compound can also be modified to have positive or negative regulatory activities. For example, a compound that binds a cis acting nucleic acid element or a nucleic acid binding factor can be modified to include a transcriptional activation domain so as to selectively activate transcription of a gene. Similarly, a compound can be modified to include a domain that would, for example, cleave a nearby nucleic acid sequence or attenuate its transcription.

Identification of cis acting nucleic acid elements also allows alteration of the genetic circuitry of a cell by genetic modification. Genetic modification can be used, for example, to enhance, reduce or alter the expression of a nucleic acid or group of nucleic acids for therapeutic purposes. For example, a normal or altered copy of one or more cis acting nucleic acid elements can be introduced at a normal location or altered location within the genome of a cell, in order to modify the regulation of a nearby nucleic acid. The cis acting nucleic acid element can be, for example, responsive to an agent such as a hormone, growth factor, metal ion or antibiotic. Following insertion, the cis acting nucleic acid element confers regulation by the agent on the nucleic acid of interest. Similarly, a strong constitutive promoter or enhancer element or elements can be inserted in close proximity to a nucleic

5 acid of interest to constitutively increase the  
expression of the nucleic acid. One or more cis acting  
10 nucleic acid elements that normally regulate a nucleic  
acid of interest can also be removed or replaced to alter  
5 the regulation of the nucleic acid.

15 Therefore, the invention provides a method of  
treating a pathological condition in an individual by  
genetic modification. The method involves contacting a  
cell of the individual with an effective amount of a  
20 targeting construct that includes a cis acting nucleic  
acid element and targeting sequences. The targeting  
sequences correspond to a sequence of a nucleic acid  
involved in the pathological condition. The targeting  
construct is taken up by the cell and the cis acting  
25 nucleic acid element is inserted by homologous  
recombination into the nucleic acid involved in the  
pathological condition so as to alter its genetic  
30 activity.

Methods of inserting, removing and replacing  
20 nucleic acid sequences at predetermined locations using  
homologous recombination are known in the art and are  
35 described, for example, in Yanez et al., Gene Therapy  
5:149-159 (1998), which is incorporated herein by  
reference. A targeting construct is prepared that  
40 25 carries a segment of nucleic acid homologous to the  
target nucleic acid as well as the desired modified  
sequences. As described above, the modified sequences  
can be, for example, a normal or altered copy of a cis  
45 acting nucleic acid element that is to be introduced into  
30 the target locus. Targeting constructs can be delivered  
to the target cells by a variety of methods known in the  
art, including, for example, electroporation,  
50 microinjection, optoporation, polybrene, DMSO, DEAE-

5 dextran, liposome formulations, gene gun, polyamidoamine  
10 dendrimers, synthetic peptides and combinations of these  
agents and methods, such that they are taken up by the  
target cells and incorporated into the target nucleic  
5 acid. Large targeting constructs for homologous  
recombination can be incorporated, for example, into  
15 plasmids, cosmids or viral vectors, such as retroviral or  
adenoviral vectors. Alternatively, chimeric DNA-RNA  
oligonucleotides or small denatured DNA fragments, which  
20 include the cis acting nucleic acid element flanked by  
short targeting sequences, can also be used to introduce  
a cis acting nucleic acid element into a cell at a  
predetermined location in the genome.

25 Homologous recombination can be practiced either  
15 *ex vivo* or *in vivo*, as needed, depending on the  
therapeutic strategy. For example, cells of a variety of  
lineages can be obtained from an individual, genetically  
30 modified *ex vivo* by insertion, deletion or replacement of  
one or more cis acting nucleic acid elements in order to  
20 enhance expression of a beneficial gene or gene product  
or reduce expression of a harmful gene or gene product,  
35 and returned to the same or an immunologically matched  
individual for therapeutic benefit. Similarly, a  
targeting construct can be used to directly contact a  
25 diseased cell within an individual, so as to be taken up  
40 by the cell and inserted into the target nucleic acid  
that is involved in the pathological condition so as to  
alter its genetic activity.

45 Cis acting nucleic acid elements can also be used  
30 to identify new genes that may be of importance in  
diagnosing and treating disease. As known in the art and  
described above, most structural and regulatory genes are  
50 characterized by the presence of cis acting nucleic acid

5 sequences either within or adjacent to the gene.  
Therefore the presence of a cis acting nucleic acid  
10 element is indicative of a nearby gene. For example, cis  
acting DNA elements can be detectably labeled and used to  
5 hybridize to genomic libraries, or libraries of  
subgenomic regions, using known methods. The genes so  
15 identified can be sequenced and identified. This  
procedure advantageously allows the simultaneous  
identification of a plurality of genes that are modulated  
10 by the same cis acting nucleic acid element or  
combination of elements.  
20

The invention also provides a method of  
determining the binding state of a nucleic acid. The  
25 method involves contacting a nucleic acid with a  
plurality of isolated cis acting nucleic acid elements  
under conditions that allow nucleic acid binding factors  
bound to the nucleic acid to bind to the isolated cis  
30 acting nucleic acid elements. The isolated cis acting  
20 nucleic acid elements that bind to the nucleic acid  
binding factors are identified, and characterize the  
binding state of the nucleic acid.  
35

Cellular nucleic acid binding factors can either  
be constitutively bound to cis acting nucleic acid  
25 elements or bind in response to appropriate extracellular  
signals. For example, nucleic acid binding factors can  
40 bind cis acting nucleic acid elements as a response to  
hormones, growth and differentiation factors, stress,  
pathological conditions, contact with neighboring cells  
45 and other such stimuli. Therefore, the binding state of  
30 a nucleic acid reflects its response to its environment  
at the time of detection.  
50  
55



5 Depending on the desired application of the  
method, a binding state can be determined for any nucleic  
acid molecule in a single cell, group of cells or tissue  
10 of interest. The nucleic acid is obtained under  
5 conditions where it remains bound to its normal nucleic  
acid binding factors. For example, a chromatin  
preparation, hnRNA preparation, mRNA preparation, or any  
15 fraction of these or other preparations described above,  
can be obtained from a single cell, group of cells or  
10 tissue. By methods described above the nucleic acid  
preparation is contacted with a plurality of isolated cis  
20 acting nucleic acid elements under conditions such that  
the nucleic acid binding factors will bind to the  
isolated cis acting nucleic acid elements. As described  
15 above, such conditions can, if desired, involve an excess  
of isolated cis acting nucleic acid elements to shift the  
25 equilibrium to favor binding to the isolated cis acting  
nucleic acid elements.

30 A plurality of isolated cis acting nucleic acid  
20 elements useful in determining the binding state of a  
nucleic acid can include any type and combination of  
isolated cis acting nucleic acid elements, as described  
35 above, such as cis acting nucleic acid elements that  
regulate a particular group of genes or are found in a  
25 particular cell type of interest. The isolated cis  
acting nucleic acid elements that bind to nucleic acid  
40 binding factors can be distinguished from unbound nucleic  
acids by any of the methods described above including,  
for example, retention on nitrocellulose, protection from  
45 30 restriction digestion, and density or size fractionation.

Methods of determining which isolated cis acting  
nucleic acid elements are bound by a nucleic acid binding  
50 factor can also be automated. Automated detection is

5 particularly advantageous in rapidly and reproducibly  
screening a large number of samples to determine their  
binding state. For example, oligonucleotides  
10 representing known cis acting nucleic acid elements can  
5 be synthesized at known positions on arrays. Those cis  
acting nucleic acid elements that are bound by nucleic  
acid binding factors have altered properties, in  
15 comparison with unbound cis acting nucleic acid elements,  
as described previously, which allow them to be detected  
20 by automated methods known in the art. The type, number,  
pattern or extent of bound cis acting nucleic acid  
elements is indicative of the binding state of the  
nucleic acid being assayed.

25 A method of the invention can be used to diagnose  
15 disease in an individual by comparing the binding state  
of nucleic acids obtained from a cell, group of cells or  
tissue of an individual suspected of having a disease  
30 with the binding state of nucleic acids obtained from  
similar cells from a normal individual. As a non-  
20 limiting example, the binding state of one or more  
nucleic acids can be used to diagnose cancer. Cancer is  
35 characterized by the enhanced expression of genes that  
promote the proliferation and metastasis of abnormal  
cells, such as growth factors, proteases, angiogenic  
25 factors, and the like. A method of the invention can be  
used, therefore, to determine whether cis acting nucleic  
40 acid elements that regulate the expression of such genes  
are bound to nucleic acid binding factors in a particular  
tissue. Cancer is also characterized by an increase in  
45 30 DNA synthesis. Therefore, a method of the invention can  
be used to determine whether cis acting nucleic acid  
elements that regulate DNA synthesis are bound in a  
particular tissue.

5 The binding state of nucleic acids can be  
determined, for example, before and after the  
administration of a therapeutic agent to monitor the  
10 consequences of therapy. For example, if a therapy is  
5 successful, the binding state of nucleic acids will more  
closely resembles the known normal binding state than the  
previous diseased state.

15 It is understood that modifications which do  
not substantially affect the activity of the various  
20 embodiments of this invention are also included within  
the definition of the invention provided herein.  
Accordingly, the following examples are intended to  
illustrate but not limit the present invention.

#### 25 EXAMPLE I

##### 15 Methods of identifying nucleic acids containing 30 a cis acting nucleic acid element and methods of isolating nucleic acid binding factors

35 This example shows a method of identifying a  
nucleic acid containing a cis acting nucleic acid  
20 element, and a method of isolating a nucleic acid binding  
factor.

40 The method is practiced by biotinylating one  
strand of a double-stranded DNA bait at the 5' end. The  
core of the double-stranded DNA bait is random over  
25 about 20 base pairs. There are restriction sites at both  
45 ends of the bait, such as Sau3A1 sites. The DNA bait  
structure is prepared by chemical synthesis of the  
biotinylated strand, and enzymatic synthesis of the

50

55

complementary strand by elongation of the appropriate primer.

The design of the bait optionally includes sequences recognized by restriction enzymes that cut at a distance from their binding site, as described previously. Nuclear proteins, optionally histone-free, are purified in bulk from cell lines or tissue nuclei (animal or plant) by standard techniques. Alternatively, chromatin, optionally histone-free, is prepared from the same sources. As a further alternative, nuclear membrane fragments are prepared by flotation in sucrose gradient in order to focus the procedure on those transcription factor complexes and other DNA binding proteins that are associated with the nuclear membrane.

Bait DNA is incubated with nuclear proteins, chromatin or nuclear membrane fragments in a buffer medium containing protease inhibitors. The bait concentration is such that there are about  $10^3$  copies of every possible 20-mer random core, which corresponds to about 5  $\mu$ g of a 50 bp bait. The incubation variables are time, temperature and ionic strength, all of which may be varied to increase specificity. The incubation mix also contains synthetic double stranded DNA of low complexity (such as polydI-polydC) to compete for proteins that have a sequence-independent affinity for DNA (non-specific binding). The mixture is then passed through a nitrocellulose filter. This step ensures that only those bait DNAs that are complexed to proteins will be retained on the filter. Recovery of the bound bait DNA is effected by mild detergent elution followed by magnetic isolation with dynabeads coated with avidin. At this stage, several procedures can be conducted in parallel:

5  
10  
15  
1) The washed beads are heated so as to denature the bound DNA and the beads are removed with a magnet, leaving the non-biotinylated strand in solution. This is amplified by PCR, using primers flanking the random sequences, one of which is derivatized in 5' with biotin. The amplified DNA is used as bait in a second round of selection, as above. The procedure may be reiterated.

20  
25  
30  
35  
2) Alternatively, washed beads are treated with restriction enzyme Sau3A1, so as to generate GATC sticky ends (which are also hemi-BamHI sites). The DNA is then ligated to an appropriate vector linearized with BamHI and dephosphorylated by alkaline phosphatase. Upon transformation into super-competent cells,  $10^4$ - $10^5$  independent clones are obtained. These are grown in bulk and their inserts further amplified by PCR as above, one of the primers being biotinylated. Alternatively, several pools of  $10^2$ - $10^3$  clones may be prepared and their inserts independently amplified. The amplified DNA is used as bait in one or more further rounds of selection, as in procedure 1).

40  
45  
50  
3) As a further alternative, nuclear membrane preparations that have been incubated with bait DNA are floated again on a sucrose gradient, and the bait DNA specifically bound to this fraction eluted by mild detergent treatment, concentrated and purified on avidin beads and submitted to amplification and rescreening as above. This ensures the selective purification and amplification of those DNA sequences that bind to nucleic acid binding factors.

55  
The specificity of the cis-element isolation procedure can be further increased by use of a

5 restriction enzyme whose recognition sequence is in the  
fixed segment of the bait DNA and whose cutting site is  
10 situated 10-20 bp to the side, designated type IIS  
restriction enzymes. Digestion of the bait DNA-nuclear  
5 protein complexes with such an enzyme selectively cleaves  
naked bait DNA and spares protein-complexed DNA. The  
cleaved DNA is not a substrate in the subsequent  
15 amplification reaction, thereby increasing the  
specificity of the procedure and selecting for  
10 protein-DNA complexes whose off-rate is slow.

20 At this stage, the sets of selected bait DNAs  
are highly enriched in sequences that are capable of  
binding nuclear proteins and nuclear membrane receptors  
effectively. An aliquot is cloned at the BamHI site of a  
25 vector and 30-40 independent clones are sequenced by  
15 priming at a distance of about 50 bp from the inserts.  
This yields a first crop of sequences among which known  
cis-elements are present, such as SP1 and AP2 sites, N  
30 and E boxes, and the like.

20 The remaining sequences in the initial set are  
35 analyzed for palindromes. Selected motifs are then  
synthesized chemically, tethered to beads and incubated  
with nuclear proteins, chromatin or nuclear membrane  
fragments. The bound proteins are then isolated  
40 25 magnetically and submitted to microsequencing. The  
N-terminal sequences are compared to the databank set of  
all known open reading frames to find whether the  
corresponding genes have previously been sequenced and  
45 what, if anything, is known about their function. If the  
30 N-terminal sequences are novel, they can be cloned and  
sequenced by established procedures.

50

55

5 To enhance the isolation of novel cis-elements,  
DNA is incubated with nuclear proteins in the presence of  
10 synthetic double stranded DNA bearing the recognition  
sequence motifs for the most prevalent and ubiquitous  
5 transcription factors.

15 The end result of this process extended to  
about  $10^3$  DNA sequences isolated from a variety of tissues  
is the isolation and identification of a set of proteins  
capable of specific binding to a large set of cis-acting  
20 nucleic acid elements. Depending on the tissue, its  
developmental stage or its pathological status,  
non-equivalent sets should be obtained, suggesting ways  
to specifically affect transcription for agriculture or  
25 biomedical applications.

15

**EXAMPLE II****Preparation of a Promoter-Library**

30

This example shows a procedure for preparing a  
library enriched for promoter sequences.

35

1. Poly-A+ mRNA is isolated from a tissue of  
20 interest.

40

2. A first strand is synthesized by reverse  
transcriptase primed by random hexamers and in the  
presence of Br-dUTP (or digoxigenin-dUTP). The use of  
random hexamers increases the probability for complete  
25 first strand synthesis extending to the mRNA cap.

45

3. The first strand of single stranded DNA is  
annealed to genomic DNA (cut with EcoRI, Hind III) under  
50 high stringency conditions.

50

55

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

4. The 3'-ends of hybrids are extended with Taq DNA polymerase in the presence of biotin-dUTP. In this step, the BrdU-labeled cDNA complementary to the transcribed sequences are extended into biotin-labeled DNA complementary to upstream (promoter) sequences.

5. The sample is digested with a restriction enzyme, such as Sau3A (creates BamHI/BglII-compatible 5'-GATC overhangs). Alternatively, other enzymes (6-cutters) could be used to make longer fragments.

6. The DNA sample is incubated consecutively with [1] anti-mouse IgG beads containing antiBrdU antibodies to purify BrdU-containing DNA (mRNA coding sequences) and [2] streptavidin beads to purify biotin-containing DNA (promoter sequences). Only DNA fragments containing both BrdU and biotin will bind to both beads. This eliminates unextended first-strand cDNA and DNA resulting from non-specific extension during step 4.

The quality of the preparation can be determined by testing for the presence of promoter sequences of known constitutively expressed genes (actin, cyclin, Ku), using primers based on GenBank sequence data.

The promoter library can be used, for example, in the following applications:

A. Use BrdU+/biotin+ fragments as templates for the preparation of random-primed 15- to 20-mer libraries.



5                   B.    Clone BrdU+/biotin+ fragments into  
BamHI-site of plasmid vector for sequencing of individual  
10                   products.

                  C.    Ligate "adapters" to Sau3A1-ends for  
5                   single-primer PCR amplification of products.

15                   D.    Use clones to generate "promoter chips."

                  Throughout this application various publications  
20                   have been referenced within parentheses. The disclosures  
of these publications in their entireties are hereby  
10                   incorporated by reference in this application in order to  
more fully describe the state of the art to which this  
25                   invention pertains.

                  Although the invention has been described with  
reference to the disclosed embodiments, those skilled in  
30                   the art will readily appreciate that the specific  
experiments detailed are only illustrative of the  
invention. It should be understood that various  
modifications can be made without departing from the  
35                   spirit of the invention. Accordingly, the invention is  
20                   limited only by the following claims.

5

10

15

20

25

30

35

40

45

50

55

## Claims

5

10

15

20

25

30

35

40

45

50

55

What is claimed is:

1. A method of identifying a nucleic acid containing a cis acting nucleic acid element, comprising:

- (a) contacting a diverse population of nucleic acid binding factors with a diverse population of isolated nucleic acid molecules under conditions that allow said nucleic acid binding factors to selectively bind said isolated nucleic acid molecules; and
- (b) identifying one or more isolated nucleic acid molecules that bind to one or more nucleic acid binding factors, said isolated nucleic acid molecules that bind to said nucleic acid binding factors being characterized as nucleic acids containing cis acting nucleic acid elements.

2. The method of claim 1, wherein said diverse population of isolated nucleic acid molecules comprises two or more different nucleic acid molecules.

3. The method of claim 1, wherein said diverse population of isolated nucleic acid molecules comprises greater than about  $10^3$  different nucleic acid molecules.

4. The method of claim 1, wherein said diverse population of isolated nucleic acid molecules comprises nucleic acids attached to a solid support.

5. The method of claim 1, wherein said diverse population of nucleic acid binding factors comprises two or more different nucleic acid binding factors.

5  
10  
6. The method of claim 1, wherein said diverse population of nucleic acid binding factors comprises greater than about  $10^3$  different nucleic acid binding factors.

5  
15  
20  
7. The method of claim 1, wherein said diverse population of nucleic acid binding factors comprises nucleic acid binding factors bound to nucleic acids selected from the group consisting of chromatin, a chromosome, a chromosome arm, a transcriptional domain, a gene family and a gene.

8. A method of isolating a nucleic acid binding factor, comprising:

- 25  
15  
30  
35  
(a) contacting a diverse population of nucleic acid binding factors with a diverse population of isolated nucleic acid molecules under conditions that allow said nucleic acid binding factors to selectively bind said isolated nucleic acid molecules; and  
(b) isolating one or more of said nucleic acid binding factors that selectively bind to one or more of said isolated nucleic acid molecules.

9. The method of claim 8, wherein said diverse population of isolated nucleic acid molecules comprises two or more different nucleic acid molecules.

40  
25  
45  
10. The method of claim 8, wherein said diverse population of isolated nucleic acid molecules comprises greater than about  $10^3$  different nucleic acid molecules.

50  
30  
11. The method of claim 8, wherein said diverse population of isolated nucleic acid molecules comprises nucleic acids attached to a solid support.

5  
10  
12. The method of claim 8, wherein said diverse population of nucleic acid binding factors comprises two or more different nucleic acid binding factors.

5  
15  
13. The method of claim 8, wherein said diverse population of nucleic acid binding factors comprises greater than about  $10^3$  different nucleic acid binding factors.

20  
10  
25  
14. The method of claim 8, wherein said diverse population of nucleic acid binding factors comprises nucleic acid binding factors bound to nucleic acids selected from the group consisting of chromatin, a chromosome, a chromosome arm, a transcriptional domain, a gene family and a gene.

15  
30  
15. A method of identifying a cis acting nucleic acid element analog, comprising:

35  
20  
(a) contacting a diverse population of nucleic acid binding factors with a diverse population of compounds under conditions that allow said compounds to selectively bind said nucleic acid binding factors;

40  
25  
(b) contacting one or more of said nucleic acid binding factors selectively bound to one or more of said bound compounds with one or more isolated nucleic acid molecules under conditions whereby one or more of said isolated nucleic acid molecules selectively displaces one or more of said selectively bound compounds from one or more of said bound nucleic acid binding factors; and

45  
30  
(c) identifying one or more of said displaced compounds, said compounds being characterized as cis acting nucleic acid element analogs.

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

16. The method of claim 15, further comprising identifying one or more of said isolated nucleic acid molecules that displaces one or more of said selectively bound compounds, said isolated nucleic acid molecule being characterized as a nucleic acid containing a cis-acting nucleic acid element.

17. The method of claim 15, wherein said diverse population of nucleic acid binding factors comprises two or more different nucleic acid binding factors.

18. The method of claim 15, wherein said diverse population of nucleic acid binding factors comprises greater than about  $10^3$  different nucleic acid binding factors.

19. The method of claim 15, wherein said diverse population of nucleic acid binding factors comprises nucleic acid binding factors bound to nucleic acids selected from the group consisting of chromatin, a chromosome, a chromosome arm, a transcriptional domain, a gene family and a gene.

20. The method of claim 15, wherein said diverse population of compounds comprises greater than about  $10^5$  different compounds.

5

21. A method of identifying a nucleic acid binding factor analog, comprising:

10

(a) contacting a diverse population of compounds with a diverse population of isolated nucleic acid molecules under conditions that allow said compounds to selectively bind said isolated nucleic acid molecules;

15

(b) contacting one or more of said isolated nucleic acid molecules selectively bound to one or more of said compounds with one or more nucleic acid binding factors under conditions whereby one or more of said nucleic acid binding factors selectively displaces one or more of said selectively bound compounds from one or more of said bound nucleic acids; and

20

25

(c) identifying one or more of said displaced compounds, said compounds being characterized as nucleic acid binding factor analogs.

30

22. The method of claim 21, further comprising identifying one or more of said nucleic acid binding factors that selectively displaces one or more of said selectively bound compounds from one or more of said bound nucleic acids.

35

23. The method of claim 21, wherein said diverse population of isolated nucleic acid molecules comprises two or more different nucleic acid molecules.

40

24. The method of claim 21, wherein said diverse population of isolated nucleic acid molecules comprises greater than about  $10^5$  different nucleic acid molecules.

45

50

25. The method of claim 21, wherein said diverse population of isolated nucleic acid molecules comprises nucleic acids attached to a solid support.

55

5 26. The method of claim 21, wherein said  
diverse population of compounds comprises greater than  
10 about  $10^5$  different compounds.

15 27. A method of identifying a compound that  
5 selectively binds a cis acting nucleic acid element,  
comprising:

15 (a) contacting a plurality of isolated nucleic  
acid molecules with a diverse population of compounds  
under conditions that allow said compounds to selectively  
20 bind said isolated nucleic acid molecules, each isolated  
nucleic acid molecule comprising one or more cis acting  
nucleic acid elements; and

25 (b) identifying one or more compounds that  
selectively bind one or more isolated nucleic acid  
15 molecules comprising a cis acting nucleic acid element.

30 28. The method of claim 27, wherein said  
diverse population of compounds comprises greater than  
about  $10^5$  different compounds.

35 29. The method of claim 27, wherein said  
20 plurality of isolated nucleic acid molecules comprises  
between about 2-5 different isolated nucleic acid  
molecules, preferably between about 6-10 different  
isolated nucleic acid molecules, more preferably between  
40 about 11-20 different isolated nucleic acid molecules,  
25 most preferably greater than about 20 different isolated  
nucleic acid molecules.

45 30. The method of claim 27, wherein said  
plurality of isolated nucleic acid molecules comprises  
nucleic acids attached to a solid support.

50

55



5  
10  
31. A method of identifying a compound that selectively displaces binding between a nucleic acid binding factor and a cis acting nucleic acid element or a nucleic acid binding factor, comprising:

5 (a) contacting a diverse population of isolated nucleic acid molecules selectively bound to nucleic acid binding factors with a diverse population of compounds under conditions that allow said compounds to selectively displace one or more of said selectively bound nucleic  
10 acid binding factors; and

20 (b) identifying one or more compounds that selectively displace one or more of said bound nucleic acid binding factors.

25  
32. The method of claim 31, wherein said one or more compounds that displace one or more of said bound nucleic acid binding factors is a cis acting nucleic acid element analog.

30  
33. The method of claim 31, wherein said one or more compounds that displace one or more of said bound nucleic acid binding factors is a nucleic acid binding factor analog.

40  
34. The method of claim 31, wherein said diverse population of compounds comprises greater than about  $10^5$  different compounds.

45  
35. The method of claim 31, wherein said diverse population of isolated nucleic acid molecules comprises two or more different nucleic acid molecules.

5 36. The method of claim 31, wherein said  
diverse population of isolated nucleic acid molecules  
10 comprises greater than about  $10^5$  different nucleic acid  
molecules.

5 37. The method of claim 31, wherein said  
15 plurality of isolated nucleic acid molecules comprises  
nucleic acids attached to a solid support.

20 38. A plurality of isolated nucleic acid  
molecules, each isolated nucleic acid molecule comprising  
10 one or more cis acting nucleic acid elements.

25 39. The plurality of claim 38, comprising  
between about 2-5 different isolated nucleic acid  
molecules, preferably between about 6-10 different  
isolated nucleic acid molecules, more preferably between  
15 about 11-20 different isolated nucleic acid molecules,  
30 most preferably greater than about 20 different isolated  
nucleic acid molecules.

35 40. The plurality of claim 38, comprising  
isolated nucleic acid molecules attached to a solid  
20 support.

40 41. A plurality of isolated nucleic acid  
molecules bound to nucleic acid binding factors, each  
isolated nucleic acid molecule comprising one or more cis  
acting nucleic acid elements.

45 25 42. The plurality of claim 41, comprising  
between about 2-5 different isolated nucleic acid  
molecules, preferably between about 6-10 different  
50 isolated nucleic acid molecules, more preferably between

about 11-20 different isolated nucleic acid molecules,  
most preferably greater than about 20 different isolated  
nucleic acid molecules.

43. The plurality of claim 41, comprising  
isolated nucleic acid molecules attached to a solid  
support.

44. A plurality of isolated cis acting nucleic  
acid element analogs.

45. The plurality of claim 44, comprising  
between about 2-5 different isolated cis acting nucleic  
acid element analogs, preferably between about 6-10  
different isolated cis acting nucleic acid element  
analog, more preferably between about 11-20 different  
isolated cis acting nucleic acid element analogs, most  
preferably greater than about 20 different isolated cis  
acting nucleic acid element analogs.

46. A plurality of isolated cis acting nucleic  
acid element analogs bound to nucleic acid binding  
factors.

47. The plurality of claim 46, comprising  
between about 2-5 different isolated cis acting nucleic  
acid element analogs, preferably between about 6-10  
different isolated cis acting nucleic acid element  
analog, more preferably between about 11-20 different  
isolated cis acting nucleic acid element analogs, most  
preferably greater than about 20 different isolated cis  
acting nucleic acid element analogs.

5 48. A plurality of isolated nucleic acid  
binding factors, said plurality comprising at least about  
10 15 different isolated nucleic acid binding factors.

15 49. The plurality of claim 48, comprising  
5 between about 16-25 different isolated nucleic acid  
binding factors, preferably between about 26-50 different  
isolated nucleic acid binding factors, more preferably  
greater than about 51 different isolated nucleic acid  
binding factors.

20 50. The plurality of claim 48, comprising  
isolated nucleic acid binding factors attached to a solid  
support.

25 51. A method of determining a binding state of  
a nucleic acid, comprising:

30 15 (a) contacting a nucleic acid with a plurality of  
isolated cis acting nucleic acid elements under  
conditions that allow nucleic acid binding factors bound  
to said nucleic acid to bind said isolated cis acting  
nucleic acid elements; and

35 20 (b) identifying said cis acting nucleic acid  
elements that bind to said nucleic acid binding factors,  
said cis acting nucleic acid elements that bind to said  
nucleic acid binding factors characterizing the binding  
40 state of said nucleic acid.

45 25 52. The method of claim 51, wherein said  
binding state is characteristic of a pathological  
condition selected from the group consisting of cancer,  
degenerative diseases, genetic disorders, immune  
disorders, bacterial infectious diseases and viral  
50 30 infectious diseases.

5

10

15

53. The method of claim 51, wherein said plurality of isolated cis acting nucleic acid elements comprises between about 2-5 different isolated nucleic acid molecules, preferably between about 6-10 different isolated nucleic acid molecules, more preferably between about 11-20 different isolated nucleic acid molecules, most preferably greater than about 20 different isolated nucleic acid molecules.

20

54. The method of claim 51, wherein said plurality of isolated cis acting nucleic acid elements comprises isolated cis acting nucleic acid elements attached to a solid support.

25

30

55. A method of treating a pathological condition in an individual, comprising administering to said individual an effective amount of one or more therapeutic agents that selectively alter the ability of one or more cis acting nucleic acid elements to regulate a genetic activity of one or more nucleic acids involved in said pathological condition.

35

56. The method of claim 55, wherein said therapeutic agent is a cis acting nucleic acid element.

40

57. The method of claim 55, wherein said therapeutic agent is a cis acting nucleic acid element analog.

45

58. The method of claim 55, wherein said therapeutic agent is a nucleic acid binding factor.

50

59. The method of claim 55, wherein said therapeutic agent is a nucleic acid binding factor analog.

55

5 60. The method of claim 55, wherein said  
therapeutic agent selectively increases a genetic  
10 activity of said one or more nucleic acids.

15 61. The method of claim 55, wherein said  
5 therapeutic agent selectively decreases a genetic  
activity of said one or more nucleic acids.

20 62. The method of claim 55, wherein said  
genetic activity is selected from the group consisting of  
nucleic acid replication, repair, packaging,  
10 modification, restriction, degradation, transcription,  
structural integrity, translation, splicing, editing,  
intracellular transport, localization and reverse  
25 transcription.

30 63. The method of claim 55, wherein said  
15 pathological condition is mediated by the dysregulation  
of one or more nucleic acids involved in said  
30 pathological condition.

35 64. The method of claim 55, wherein said  
pathological condition is mediated by a pathogen.

40 65. The method of claim 55, wherein said  
20 pathological condition is selected from the group  
consisting of cancer, degenerative diseases, genetic  
disorders, immune disorders, bacterial infectious  
45 diseases and viral infectious diseases.

45

50

55

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

66. A method of treating a pathological condition in an individual, comprising contacting a cell of said individual with an effective amount of a targeting construct comprising a cis acting nucleic acid element and targeting sequences, said targeting sequences corresponding to a nucleic acid involved in said pathological condition, said contacting being of sufficient duration so as to allow said targeting construct to be taken up by said cell and said cis acting nucleic acid element to be inserted by homologous recombination into said nucleic acid involved in said pathological condition, said inserted cis acting nucleic acid element having the effect of altering a genetic activity of said nucleic acid in said cell.

67. The method of claim 66, wherein said targeting construct contacts a cell in an individual.

68. The method of claim 66, wherein said targeting construct contacts a cell ex vivo and said cell is returned to said individual.

69. The method of claim 66, wherein said genetic activity is selected from the group consisting of nucleic acid replication, repair, packaging, modification, restriction, degradation, transcription, structural integrity, translation, splicing, editing, intracellular transport, localization and reverse transcription.

70. The method of claim 69, wherein said pathological condition is selected from the group consisting of cancer, degenerative diseases, genetic disorders, immune disorders, bacterial infectious diseases and viral infectious diseases.